ALPHA-DIVERSITY PROCESSES AND NORMALIZED INVERSE-GAUSSIAN DIFFUSIONS

By Matteo Ruggiero[1], Stephen G. Walker and Stefano Favaro[1]

University of Torino, University of Kent and University of Torino

The infinitely-many-neutral-alleles model has recently been extended to a class of diffusion processes associated with Gibbs partitions of two-parameter Poisson–Dirichlet type. This paper introduces a family of infinite-dimensional diffusions associated with a different subclass of Gibbs partitions, induced by normalized inverse-Gaussian random probability measures. Such diffusions describe the evolution of the frequencies of infinitely-many types together with the dynamics of the time-varying mutation rate, which is driven by an α-diversity diffusion. Constructed as a dynamic version, relative to this framework, of the corresponding notion for Gibbs partitions, the latter is explicitly derived from an underlying population model and shown to coincide, in a special case, with the diffusion approximation of a critical Galton–Watson branching process. The class of infinite-dimensional processes is characterized in terms of its infinitesimal generator on an appropriate domain, and shown to be the limit in distribution of a certain sequence of Feller diffusions with finitely-many types. Moreover, a discrete representation is provided by means of appropriately transformed Moran-type particle processes, where the particles are samples from a normalized inverse-Gaussian random probability measure. The relationship between the limit diffusion and the two-parameter model is also discussed.

1. Introduction. Considerable attention has been devoted recently to a class of diffusion processes which extends the infinitely-many-neutral-alleles model to the case of two parameters. This family takes values in the space

$$(1.1) \qquad \overline{\nabla}_\infty = \left\{ z = (z_1, z_2, \ldots) \colon z_1 \geq z_2 \geq \cdots \geq 0, \sum_{i=1}^\infty z_i \leq 1 \right\},$$

1

namely, the closure in $[0,1]^\infty$ of the infinite-dimensional ordered simplex, and is characterized, for constants $0 \le \alpha < 1$ and $\theta > -\alpha$, by the second order differential operator

$$(1.2) \qquad \mathcal{L}^{\theta,\alpha} = \frac{1}{2} \sum_{i,j=1}^\infty z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} - \frac{1}{2} \sum_{i=1}^\infty (\theta z_i + \alpha) \frac{\partial}{\partial z_i}$$

acting on a certain dense sub-algebra of the space $C(\overline{\nabla}_\infty)$ of continuous functions on $\overline{\nabla}_\infty$ (throughout the paper δ_{ij} denotes Kronecker delta). The diffusion with operator (1.2) describes the evolution of the allelic frequencies at a particular locus in a large population subject to random genetic drift and mutation, where mutation is jointly driven by the parameters (θ, α). Ethier and Kurtz (1981) characterized the corresponding process when $\alpha = 0$, whereas the two-parameter family was introduced by Petrov (2009) and further investigated by Ruggiero and Walker (2009) and Feng and Sun (2010). The latter is known to be stationary, reversible and ergodic with respect to the Poisson–Dirichlet distribution with parameters (θ, α). This was introduced by Pitman (1995) [see also Pitman (1996) and Pitman and Yor (1997)] and extends the Poisson–Dirichlet distribution of Kingman (1975) as follows. Consider a random sequence $(V_1, V_2, \ldots)$ obtained by means of the so-called stick-breaking scheme

$$(1.3) \quad V_1 = W_1, \qquad V_n = W_n \prod_{i=1}^{n-1} (1 - W_i), \qquad W_i \overset{\mathrm{ind}}{\sim} \mathrm{Beta}(1 - \alpha, \theta + i\alpha),$$

where $0 \le \alpha < 1$ and $\theta > -\alpha$. The vector $(V_1, V_2, \ldots)$ is said to have the GEM distribution with parameters (θ, α), while the vector of descending order statistics $(V_{(1)}, V_{(2)}, \ldots)$ is said to have the Poisson–Dirichlet distribution with parameters (θ, α). The latter is also the law of the ranked frequencies of an infinite partition induced by a two-parameter Poisson–Dirichlet random probability measure, which generalizes the Dirichlet process introduced by Ferguson (1973). Two-parameter Poisson–Dirichlet models have found applications in several fields. See, for example, the monographs by Bertoin (2006) for fragmentation and coalescent theory, Pitman (2006) for excursion theory and combinatorics, Teh and Jordan (2010) for machine learning, Lijoi and Prünster (2010) for Bayesian inference and Feng (2010) for population genetics. See also Bertoin (2008), Handa (2009) and Favaro et al. (2009).

The Poisson–Dirichlet distribution and its two parameter extension in turn belong to a larger class of random discrete distributions induced by infinite partitions of Gibbs type. These were introduced by Gnedin and Pitman (2005), and applications include fragmentation and coalescent theory [Bertoin (2006), McCullagh, Pitman and Winkel (2008), Goldschmidt, Martin and Spanò (2008)], excursion theory [Pitman (2003)], statistical physics

[Berestycki and Pitman (2007)] and Bayesian nonparametric inference [Lijoi, Mena and Prünster (2005, 2007a, 2007b), Lijoi, Prünster and Walker (2008a)]. See Pitman (2006) for a comprehensive account. See also Griffiths and Spanò (2007), Lijoi, Prünster and Walker (2008b) and Ho, James and Lau (2007).

This paper introduces a class of infinite-dimensional diffusions associated with a different subclass of Gibbs-type partitions, induced by normalized inverse-Gaussian random probability measures. Such discrete distributions, recently investigated by Lijoi, Mena and Prünster (2005), are special cases of generalized gamma processes [Pitman (2003), Lijoi, Mena and Prünster (2007b)], and their intersection with two-parameter Poisson–Dirichlet models is given by the sole case $(\theta, \alpha) = (0, 1/2)$, which corresponds to a normalized stable process with parameter $1/2$ [Kingman (1975)]. The class of diffusions studied in this paper is characterized in terms of the second order differential operator

$$
\begin{aligned}
\mathcal{A} = {} & \frac{\beta}{s} \frac{\partial}{\partial s} + \frac{1}{2} s \frac{\partial^2}{\partial s^2} + \frac{1}{2} \sum_{i,j=1}^{\infty} z_i (\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \, \partial z_j} \\
& - \frac{1}{2} \sum_{i=1}^{\infty} \left(\frac{\beta}{s} z_i + \alpha \right) \frac{\partial}{\partial z_i}
\end{aligned}
$$

(1.4)

acting on a dense sub-algebra of $C_0([0, \infty) \times \overline{\nabla}_\infty)$, the space of continuous functions on $[0, \infty) \times \overline{\nabla}_\infty$ vanishing at infinity, for parameters (β, α), with $\beta = a\tau^\alpha/\alpha$, $a > 0$, $\tau > 0$ and $\alpha = 1/2$. By comparison with (1.2), it can be seen that the last two terms of (1.4) describe the time evolution of the frequencies of infinitely-many types. Common features between (1.2) and (1.4) are the variance–covariance terms $z_i(\delta_{ij} - z_j)$ and the structure of the drift or mutation terms $-[(\beta/s)z_i + \alpha]$. The distinctive feature with respect to (1.2) is given by the fact that the positive coefficient $\theta_t = \beta/S_t$ varies in time, and is driven by what is termed here α-diversity diffusion, whose operator is given by the first two terms of (1.4). Equivalently, S_t follows the stochastic differential equation

$$
\text{(1.5)} \qquad \mathrm{d}S_t = \frac{\beta}{S_t} \, \mathrm{d}t + \sqrt{S_t} \, \mathrm{d}B_t, \qquad S_t \in [0, \infty),
$$

where B_t is a standard Brownian motion. This can be seen as a particular instance of a continuous-time analog of the notion of α-diversity, introduced by Pitman (2003) for Poisson–Kingman models, which include Gibbs-type partitions. An exchangeable random partition of $\mathbb{N}$ is said to have α-diversity S if and only if there exists a random variable S, with $0 < S < \infty$ almost surely, such that, as $n \to \infty$,

$$
\text{(1.6)} \qquad K_n/n^\alpha \to S \qquad \text{a.s.,}
$$

where K_n is the number of classes of the partition of $\{1, \ldots, n\}$. The connection between (1.5) and (1.6) will become clear in Section 4, where the α-diversity diffusion will be explicitly derived.

It is to be noted that (1.2) is not a special case of (1.4). Indeed, the only way of making $\theta_t = \beta/S_t$ constant is to impose null drift and volatility in (1.5), which implies $\theta_t \equiv 0$. Hence, consistently with the above recalled relation between normalized inverse-Gaussian and Poisson–Dirichlet random measures, (1.2) and (1.4) share only the case $(\theta_t, \alpha) \equiv (0, 1/2)$. Nonetheless, an interesting connection between these classes of diffusions can be stated. In particular, it will be shown that performing the same conditioning operation in a pre-limit particle construction of normalized inverse-Gaussian diffusions yields a particular instance of the two-parameter model.

The paper is organized as follows. Section 2 recalls all relevant definitions, among which are Gibbs-type partitions, the associated generalized Pólya-urn scheme and random probability measures of generalized gamma and normalized inverse-Gaussian types. Section 3 derives some new results on generalized gamma processes which are crucial for the construction. These are concerned with the convergence of the number of species represented only once in the observed sample and with the second order approximation of the weights of the generalized Pólya-urn scheme associated with normalized inverse-Gaussian processes. In Section 4, by postulating simple population dynamics underlying the time change of the species frequencies, we derive the α-diversity diffusion for the normalized inverse-Gaussian case, by means of a time-varying analog of (1.6) with the limit intended in distribution, and highlight its main properties. In Section 5 normalized inverse-Gaussian diffusions are characterized in terms of the operator (1.4), whose closure is shown to generate a Feller semigroup on $C_0([0, \infty) \times \overline{\nabla}_\infty)$, and the associated family of processes is shown to be the limit in distribution of certain Feller diffusions with finitely-many types. Section 6 provides a discrete representation of normalized inverse-Gaussian diffusions, which are obtained as limits in distribution of certain appropriately transformed Moran-type particle processes which model individuals explicitly, jointly with the varying population heterogeneity. Finally, Section 7 shows that conditioning on the α-diversity process to be constant, that is, $S_t \equiv s$, in a pre-limit version of the particle construction yields, in the limit, the two-parameter model (1.2) with $(\theta, \alpha) = (s^2/4, 1/2)$.

2. Preliminaries. The Poisson–Dirichlet distribution and its two parameter extension belong to the class of random discrete distributions induced by infinite partitions of Gibbs type, introduced by Gnedin and Pitman (2005). An exchangeable random partition of the set of natural numbers is said to have Gibbs form if for any $1 \le k \le n$ and any $(n_1, \ldots, n_k)$ such that $n_j \in \{1, \ldots, n\}$, for $j = 1, \ldots, k$, and $\sum_{j=1}^k n_j = n$, the law $\Pi_k^{(n)}$ of the parti-

tion $(n_1, \ldots, n_k)$ can be written as the product

$$(2.1) \qquad \Pi_k^{(n)}(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^k (1 - \alpha)_{n_j - 1}.$$

Here $0 \leq \alpha < 1$,

$$(2.2) \qquad (a)_0 = 1, \qquad (a)_m = a(a+1)\cdots(a+m-1), \qquad m > 1,$$

is the Pochhammer symbol and the coefficients $\{V_{n,k} : k = 1, \ldots, n; n \geq 1\}$ satisfy the recursive equation

$$(2.3) \qquad V_{n,k} = (n - \alpha k)V_{n+1,k} + V_{n+1,k+1}.$$

The law of an exchangeable partition is uniquely determined by the function $\Pi_k^{(n)}(n_1, \ldots, n_k)$, called the *exchangeable partition probability function*, which satisfies certain consistency conditions, which imply invariance under permutations of $\{1, \ldots, n\}$ and coherent marginalization over the $(n+1)$th item. Hence, the law of a Gibbs partition is uniquely determined by the family $\{V_{n,k} : k = 1, \ldots, n; n \geq 1\}$. Furthermore, a random discrete probability measure governing a sequence of exchangeable observations is said to be of a Gibbs type if it induces a partition which can be expressed as in (2.1). These have associated predictive distributions which generalize the Blackwell and MacQueen (1973) Pólya-urn scheme to

$$(2.4) \qquad \begin{aligned} &\mathbb{P}\{X_{n+1} \in \cdot | X_1, \ldots, X_n\} \\ &\qquad = g_0(n, K_n)\nu_0(\cdot) + g_1(n, K_n) \sum_{j=1}^{K_n} (n_j - \alpha)\delta_{X_j^*}(\cdot), \end{aligned}$$

where ν_0 is a nonatomic probability measure, $X_1^*, \ldots, X_{K_n}^*$ are the K_n distinct values observed in $X_1, \ldots, X_n$ with absolute frequencies $n_1, \ldots, n_{K_n}$, and the coefficients g_0 and g_1 are given by

$$(2.5) \qquad g_0(n,k) = \frac{V_{n+1,k+1}}{V_{n,k}}, \qquad g_1(n,k) = \frac{V_{n+1,k}}{V_{n,k}}$$

with $\{V_{n,k} : k = 1, \ldots, n; n \geq 1\}$ as above. It will be of later use to note that integrating both sides of (2.4) yields

$$(2.6) \qquad g_0(n, K_n) + (n - \alpha K_n)g_1(n, K_n) = 1,$$

also obtained from (2.3) and (2.5). Examples of Gibbs-type random probability measures are the Dirichlet process [Ferguson (1973)], obtained, for example, from (2.4) by setting $\theta > 0$ and $\alpha = 0$ in

$$(2.7) \qquad g_0(n,k) = \frac{\theta + \alpha k}{\theta + n}, \qquad g_1(n,k) = \frac{1}{\theta + n},$$

the two-parameter Poisson–Dirichlet process [Pitman (1995, 1996)], obtained from (2.7) with $0 < \alpha < 1$ and $\theta > -\alpha$, the normalized stable process [King-

man (1975)], obtained from (2.7) with $0 < \alpha < 1$ and $\theta = 0$, the normalized inverse-Gaussian process [Lijoi, Mena and Prünster (2005)] and the normalized generalized gamma process [Pitman (2003), Lijoi, Mena and Prünster (2007b)]. See also Gnedin (2010) for a Gibbs-type model with finitely-many types.

The normalized generalized gamma process is a random probability measure with representation

$$(2.8) \qquad \mu = \sum_{i=1}^{\infty} P_i \delta_{X_i},$$

whose weights $\{P_i, i \in \mathbb{N}\}$ are obtained by means of the normalization

$$(2.9) \qquad P_i = \frac{J_i}{\sum_{k=1}^{\infty} J_k},$$

where $\{J_i, i \in \mathbb{N}\}$ are the points of a generalized gamma process, introduced by Brix (1999). This is obtained from a Poisson random process on $[0, \infty)$ with mean intensity

$$\lambda(\mathrm{d}s) = \frac{1}{\Gamma(1 - \alpha)} \exp(-\tau s) s^{-(1+\alpha)} \, \mathrm{d}s, \qquad s \geq 0,$$

with $0 < \alpha < 1$ and $\tau \geq 0$, so that if $N(A)$ is the number of J_i's which fall in $A \in \mathscr{B}([0, \infty))$, then $N(A)$ is Poisson distributed with mean $\lambda(A)$. Lijoi, Mena and Prünster (2007b) showed that a generalized gamma random measure defined via (2.8) and (2.9), denoted by $\mathrm{GG}(\beta, \alpha)$, where $\beta = a\tau^{\alpha}/\alpha$ with $a > 0$ and $\tau > 0$, induces a random partition of Gibbs type with coefficients $g_0(n, K_n)$ and $g_1(n, K_n)$ in (2.4) given by

$$(2.10) \qquad \begin{aligned} g_0(n, k) &= \frac{\alpha \sum_{i=0}^{n} \binom{n}{i} (-1)^i \beta^{i/\alpha} \Gamma(k + 1 - i/\alpha; \beta)}{n \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\alpha} \Gamma(k - i/\alpha; \beta)}, \\ g_1(n, k) &= \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i \beta^{i/\alpha} \Gamma(k - i/\alpha; \beta)}{n \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\alpha} \Gamma(k - i/\alpha; \beta)}, \end{aligned}$$

where $\Gamma(c; x)$ denotes the upper incomplete gamma function

$$(2.11) \qquad \Gamma(c; x) = \int_x^{\infty} s^{c-1} \exp(-s) \, \mathrm{d}s.$$

Special cases of a generalized gamma process with parameters (β, α) are the Dirichlet process, obtained by letting $\tau = 1$ and $\alpha \to 0$, the normalized stable process, obtained by setting $\beta = 0$, and the normalized inverse-Gaussian process, obtained by setting $\alpha = 1/2$.

We conclude the section with a brief discussion of the interpretation of α in the context of species sampling with Gibbs-type partitions. Suppose

K_n different species have been observed in the first n samples from (2.4). The probability that a further sample is an already observed species is $g_1(n, K_n)(n - \alpha K_n)$, but this mass is not allocated proportionally to the current frequencies. The ratio of probabilities assigned to any pair of species (i, j) is

$$r_{i,j} = \frac{n_i - \alpha}{n_j - \alpha}.$$

When $\alpha \to 0$, the probability of sampling species i is proportional to the absolute frequency n_i. However, since for $n_i > n_j$, $(n_i - \alpha)/(n_j - \alpha)$ is increasing in α, a value of $\alpha > 0$ reallocates some probability mass from type j to type i, so that, for example, for $n_i = 2$ and $n_j = 1$ we have $r_{i,j} = 2, 3, 5$ for $\alpha = 0, 0.5, 0.75$, respectively. Thus, α has a reinforcement effect on those species that have higher frequency. See Lijoi, Mena and Prünster (2007b) for a more detailed treatment of this aspect.

3. Some results on generalized gamma random measures. In this section we investigate some properties of generalized gamma random measures which will be used in the subsequent constructions. In particular, these regard the convergence of the number of species represented only once in the observed sample, and the second order approximation of the weights of the generalized Pólya-urn scheme associated with normalized inverse-Gaussian processes.

Let $X_1, \ldots, X_n$ be an n-sized sample drawn from a generalized gamma process with parameters (β, α), let K_n denote the number of distinct species observed in the sample, and let $\mathbf{N}_n := (N_1, \ldots, N_{K_n})$ denote the vector of absolute frequencies associated with each observed species. The probability distribution of the random variable $(K_n, \mathbf{N}_n)$, for any $n \geq 1$, $k = 1, \ldots, n$ and frequencies $(n_1, \ldots, n_k)$ such that $\sum_{i=1}^{k} n_i = n$, is provided by Lijoi, Mena and Prünster (2007b) and coincides with

$$
\begin{aligned}
(3.1) \quad & \mathbb{P}(K_n = k, \mathbf{N}_n = (n_1, \ldots, n_{K_n})) \\
& = \frac{\alpha^{k-1} e^{\beta} \prod_{j=1}^{k} (1 - \alpha)_{(n_j - 1)}}{\Gamma(n)} \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha} \Gamma\left(k - \frac{s}{\alpha}; \beta\right),
\end{aligned}
$$

where $(1 - \alpha)_{(n_j - 1)}$ and $\Gamma(k - s/\alpha; \beta)$ are as in (2.2) and (2.11), respectively. Denote now by $M_{j,n}$ the number of species represented j times in the sample. Then from equation 1.52 in Pitman (2006) it follows that the distribution of $\mathbf{M}_n := (M_{1,n}, \ldots, M_{n,n})$ is given by

$$
\begin{aligned}
& \mathbb{P}(\mathbf{M}_n = (m_{1,n}, \ldots, m_{n,n})) \\
(3.2) \quad & = n! \frac{\alpha^{k-1} e^{\beta}}{\Gamma(n)} \prod_{j=1}^{n} \left(\frac{(1 - \alpha)_{(j-1)}}{j!}\right)^{m_{j,n}}
\end{aligned}
$$

$$\times \frac{1}{m_{j,n}!} \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha} \Gamma\left(k - \frac{s}{\alpha}; \beta\right)$$

for any $n \geq 1$, $k = 1, \ldots, n$ and vector $(m_{1,n}, \ldots, m_{n,n}) \in \mathcal{M}_{n,k}$, where

$$\mathcal{M}_{n,k} = \left\{ (m_{1,n}, \ldots, m_{n,n}) : m_{i,n} \geq 1, \sum_{i=1}^{n} m_{i,n} = k, \sum_{i=1}^{n} i m_{i,n} = n \right\}.$$

The following proposition identifies the speed of convergence of the number of species represented once in the sample. Denote by $\mathscr{C}(n, k, \alpha)$ the generalized factorial coefficient

$$(3.3) \qquad \mathscr{C}(n, k, \alpha) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-j\alpha)_n,$$

where $\mathscr{C}(0, 0, \alpha) = 1$ and $\mathscr{C}(n, 0, \alpha) = 0$. See Charalambides [(2005), Chapter 2] for a complete account.

PROPOSITION 3.1. *Under the normalized generalized gamma process with parameters (β, α), one has*

$$\mathbb{P}(M_{1,n} = m_{1,n})$$

$$= \frac{\alpha^{m_{1,n}-1} e^\beta}{\Gamma(n) m_{1,n}!} \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha}$$

$$(3.4) \qquad \times \sum_{j=0}^{n-m_{1,n}} (-\alpha)^j \frac{(n - m_{1,n} - j + 1)_{(m_{1,n}+j)}}{j!}$$

$$\times \sum_{k=0}^{n-m_{1,n}-j} \mathscr{C}(n - m_{1,n} - j, k, \alpha)$$

$$\times \Gamma\left(k + m_{1,n} + j - \frac{s}{\alpha}; \beta\right).$$

Moreover,

$$(3.5) \qquad \frac{M_{1,n}}{n^\alpha} \to \alpha S_\alpha \qquad a.s.,$$

where S_α is a strictly positive and almost surely finite random variable with density function

$$g_{S_\alpha}(s; \alpha, \beta) = e^{\beta - (\beta/s)^{1/\alpha}} \frac{f(s^{-1/\alpha}; \alpha)}{\alpha s^{1+1/\alpha}}$$

with $f(\cdot; \alpha)$ being the density of a positive stable random variable with parameter α.

PROOF. Denote $(x)_{[m]} = x(x-1)\cdots(x-m+1)$. From (3.2), for any $r \geq 1$ one has

$$\mathbb{E}[(M_{1,n})_{[r]}]$$

$$= \sum_{k=1}^{n} \sum_{\mathcal{M}_{n,k}} n! \frac{\alpha^{k-1} e^{\beta}}{\Gamma(n)} \prod_{j=1}^{n} \left(\frac{(1-\alpha)_{(j-1)}}{j!} \right)^{m_{j,n}} \frac{1}{m_{j,n}!} (m_{1,n})_{[r]}$$

$$\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha} \Gamma\left(k - \frac{s}{\alpha}; \beta \right)$$

$$= \sum_{k=1}^{n} \sum_{\mathcal{M}_{n,k}} n! \frac{\alpha^{k-1} e^{\beta}}{\Gamma(n)(m_{1,n}-r)!} \prod_{j=2}^{n} \left(\frac{(1-\alpha)_{(j-1)}}{j!} \right)^{m_{j,n}} \frac{1}{m_{j,n}!}$$

$$\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha} \Gamma\left(k - \frac{s}{\alpha}; \beta \right)$$

$$= \sum_{k=1}^{n} \sum_{\mathcal{M}_{n-r,k-r}} n! \frac{\alpha^{k-1} e^{\beta}}{\Gamma(n)} \prod_{j=1}^{n} \left(\frac{(1-\alpha)_{(j-1)}}{j!} \right)^{m_{j,n}} \frac{1}{m_{j,n}!}$$

$$\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha} \Gamma\left(k - \frac{s}{\alpha}; \beta \right).$$

In particular, by using the definition of generalized factorial coefficient in terms of sum over the set of partitions $\mathcal{M}_{n,k}$ [see Charalambides (2005), equation 2.62], we have

$$\sum_{M_{n-r,k-r}} \prod_{j=1}^{n} \left(\frac{(1-\alpha)_{(j-1)}}{j!} \right)^{m_{j,n}} \frac{1}{m_{j,n}!}$$

$$= \frac{(n)_{[r]}}{n! \alpha^{k-r}} \mathscr{C}(n-r, k-r, \alpha).$$

Therefore, we obtain

$$\mathbb{E}[(M_{1,n})_{[r]}]$$

(3.6)
$$= \sum_{k=1}^{n} \frac{\alpha^{r-1} (n)_{[r]} e^{\beta}}{\Gamma(n)} \mathscr{C}(n-r, k-r, \alpha)$$

$$\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha} \Gamma\left(k - \frac{s}{\alpha}; \beta \right).$$

In order to obtain the distribution of the random variable $M_{1,n}$, we can make use of the probability generating function of $M_{1,n}$, denoted $G_{(M_{1,n})}(t)$. From

(3.6) we have

$$
G_{(M_{1,n})}(t) = \sum_{r=0}^{\infty} \frac{\alpha^{r-1} e^{\beta}(n)_{[r]}}{\Gamma(n)}
$$

$$
\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha}
$$

$$
\times \sum_{k=0}^{n} \mathscr{C}(n-r,k,\alpha) \Gamma\left(k+r-\frac{s}{\alpha};\beta\right) \frac{(t-1)^r}{r!}.
$$

Therefore, the distribution of $M_{1,n}$ is given by

$$
\mathbb{P}(M_{1,n} = m_{1,n})
$$

$$
= \frac{1}{m_{1,n}!} \sum_{j=0}^{\infty} \frac{\alpha^{m_{1,n}+j-1} e^{\beta}(n)_{[m_{1,n}+j]}}{\Gamma(n)}
$$

$$
\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha}
$$

$$
\times \sum_{k=0}^{n} \mathscr{C}(n-m_{1,n}-j,k,\alpha)
$$

$$
\times \Gamma\left(k+m_{1,n}+j-\frac{s}{\alpha};\beta\right) \frac{d^{m_{1,n}}}{dt^{m_{1,n}}} \frac{(t-1)^{m_{1,n}+j}}{(m_{1,n}+j)!}\bigg|_{t=0}
$$

$$
= \frac{1}{m_{1,n}!} \sum_{j=0}^{\infty} \frac{\alpha^{m_{1,n}+j-1} e^{\beta}(n)_{[m_{1,n}+j]}}{\Gamma(n)}
$$

$$
\times \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha}
$$

$$
\times \sum_{k=1}^{n-m_{1,n}-j} \mathscr{C}(n-m_{1,n}-j,k,\alpha)
$$

$$
\times \Gamma\left(k+m_{1,n}+j-\frac{s}{\alpha};\beta\right) \frac{(-1)^j}{j!}
$$

$$
= \frac{\alpha^{m_{1,n}-1} e^{\beta}}{\Gamma(n)m_{1,n}!} \sum_{s=0}^{n-1} \binom{n-1}{s} (-1)^s \beta^{s/\alpha}
$$

$$
\times \sum_{j=0}^{n-m_{1,n}} (-\alpha)^j \frac{(n-m_{1,n}-j+1)_{(m_{1,n}+j)}}{j!}
$$

$$\times \sum_{k=0}^{n-m_{1,n}-j} \mathscr{C}(n - m_{1,n} - j, k, \alpha)$$

$$\times \Gamma\left(k + m_{1,n} + j - \frac{s}{\alpha}; \beta\right),$$

where the last identity is due to the fact that $\mathscr{C}(n, k, \alpha) = 0$ for any $k > n$. Proposition 3 in Lijoi, Mena and Prünster (2007b) shows that

$$(3.7) \qquad\qquad\qquad K_n/n^\alpha \to S_\alpha$$

almost surely, where S_α is an almost surely positive and finite random variable with density function

$$g_{S_\alpha}(s; \alpha, \beta) = e^{\beta - (\beta/s)^{1/\alpha}} \frac{f(s^{-1/\alpha}; \alpha)}{\alpha s^{1+1/\alpha}}$$

with $f(\cdot; \alpha)$ being the density function of a positive stable random variable with parameter α. In other terms, according to Definition 3.10 in Pitman (2006), an exchangeable partition of $\mathbb{N}$ having EPPF (3.1) has α-diversity S_α. A simple application of Lemma 3.11 in Pitman (2006) leads to (3.5). $\square$

A second aspect of generalized gamma random measures we need to address for later use is the approximate behavior of the coefficients in the generalized Pólya urn (2.4). It is well known that the first order behavior of (2.10) is that of a normalized stable process, that is,

$$(3.8) \qquad\qquad g_0(n, k) \approx \alpha k/n, \qquad g_1(n, k) \approx 1/n,$$

also implied by the next result. However, it turns out that for the definition of the diffusion processes which are the object of the next two sections, it is crucial to know the second order approximation. The following proposition, whose proof is deferred to the Appendix, identifies such behavior for the normalized inverse-Gaussian case $\alpha = 1/2$.

PROPOSITION 3.2.   *Let $g_0(n, k)$ and $g_1(n, k)$ be as in (2.10). When $\alpha = 1/2$,*

$$g_0(n, k) = \frac{\alpha k}{n} + \frac{\beta/s_n}{n} + o(n^{-1})$$

and

$$(3.9) \qquad\qquad g_1(n, k) = \frac{1}{n} - \frac{\beta/s_n}{n^2} + o(n^{-2}),$$

where $s_n = k/n^\alpha$ and $\beta = a\tau^\alpha/\alpha$.

4. Alpha-diversity processes.   Making use of the results of the previous section, here we construct a one-dimensional diffusion process which can be seen as a dynamic version of the notion of α-diversity, recalled in (1.6), relative to the case of normalized inverse-Gaussian random probability mea-

sures. Such diffusion, which will be crucial for the construction of Section 5, is obtained as weak limit of an appropriately rescaled random walk on the integers, whose dynamics are driven by an underlying population process. This is briefly outlined here and will be formalized in Section 6. Consider n particles, denoted $x^{(n)} = (x_1, \ldots, x_n)$ with $x_i \in \mathbb{X}$ for each i, where $\mathbb{X}$ is a Polish space, and denote by $K_n = K_n(x^{(n)})$ the number of distinct values observed in $(x_1, \ldots, x_n)$. Let the vector $(x_1, \ldots, x_n)$ be updated at discrete times by replacing a uniformly chosen coordinate. Conditionally on $K_n(x^{(n)}) = k$, the incoming particle will be a copy of one still in the vector, after the removal, with probability $g_1(n-1, k_r)$, and will be a new value with probability $g_0(n-1, k_r)$, where $g_1(n-1, k)$ and $g_0(n-1, k)$ are as in (2.10) and k_r is the value of k after the removal. Denote by $\{K_n(m), m \in \mathbb{N}_0\}$ the chain which keeps track of the number of distinct types in $(x_1, \ldots, x_n)$. Then, letting $m_{1,n}$ be the number of clusters of size one in $(x_1, \ldots, x_n)$, which, by means of (3.5) and (3.7) is approximately αk for large n, the transition probabilities for $K_n(m)$,

$$p(k, k') = \mathbb{P}\{K_n(m+1) = k' | K_n(m) = k\}$$

are asymptotically equivalent to

$$(4.1) \quad p(k, k') = \begin{cases} \left(1 - \dfrac{\alpha k}{n}\right) g_0(n-1, k), & \text{if } 1 \leq k < n, k' = k+1, \\ \dfrac{\alpha k}{n} g_1(n-1, k-1)(n-1-\alpha(k-1)), & \\ & \text{if } 1 < k \leq n, k' = k-1, \\ 1 - p(k, k+1) - p(k, k-1), & \text{if } k' = k, \\ 0, & \text{else} \end{cases}$$

for $1 \leq k \leq n$. That is, with probability $m_{1,n}/n \approx \alpha k/n$ a cluster of size one is selected and removed, with probability $g_0(n-1, k)$ a new species appears and with probability $g_1(n-1, k)(n-1-\alpha(k-1))$ a survivor has an offspring. Note that $k = 1$ and $k = n$ are set to be barriers, to render the fact that $m_{1,n}$ equals 0 and n when k equals 1 and n, respectively.

The following theorem finds the conditions under which the rescaled chain $K_n(m)/n^\alpha$ converges to a diffusion process on $[0, \infty)$. Here we provide a sketch of the proof with the aim of favoring the intuition. The formalization of the result is contained in the proof of Theorem 6.1, while that of the fact that the limiting diffusion is well defined, that is, the corresponding operator generates a Feller semigroup on an appropriate subspace of $C([0, \infty))$, is provided in Corollary 4.1 below.

Throughout the paper $C_B(A)$ denotes the space of continuous functions from A to B, while $X_n \Rightarrow X$ denotes convergence in distribution.

THEOREM 4.1. *Let $\{K_n(m), m \in \mathbb{N}_0\}$ be a Markov chain with transition probabilities as in (4.1) determined by a generalized gamma process*

with $\beta \geq 0$ and $\alpha = 1/2$, and define $\{\tilde{K}_n(t), t \geq 0\}$ to be such that $\tilde{K}_n(t) = K_n(\lfloor n^{3/2} t \rfloor)/n^\alpha$. Let also $\{S_t, t \geq 0\}$ be a diffusion process driven by the stochastic differential equation

$$(4.2) \qquad dS_t = \frac{\beta}{S_t} dt + \sqrt{S_t} dB_t, \qquad S_t \geq 0,$$

where B_t is a standard Brownian motion. If $\tilde{K}_n(0) \Rightarrow S_0$, then

$$(4.3) \quad \{\tilde{K}_n(t), t \geq 0\} \Rightarrow \{S_t, t \geq 0\} \qquad in \ C_{[0,\infty)}([0,\infty)) \ as \ n \to \infty.$$

PROOF. Let $\alpha = 1/2$. From Proposition 3.2 we can write (4.1) as follows (for ease of presentation we use n and k in place of $n-1$ and $k-1$ since it is asymptotically equivalent):

$$p(k, k') = \begin{cases} \left(1 - \dfrac{\alpha k}{n}\right)\left(\dfrac{\alpha k}{n} + \dfrac{\beta/s_n}{n}\right) + o(n^{-1}), & \text{if } 1 \leq k < n, k' = k+1, \\[2mm] \dfrac{\alpha k}{n}\left(\dfrac{1}{n} - \dfrac{\beta/s_n}{n^2}\right)(n - \alpha k) + o(n^{-3/2}), & \text{if } 1 < k \leq n, k' = k-1, \\[2mm] 1 - p(k, k+1) - p(k, k-1) + o(n^{-1}), & \text{if } k' = k, \\[1mm] 0, & \text{else.} \end{cases}$$

The conditional expected increment of the process $\{K_n(m)/n^\alpha, m \in \mathbb{N}_0\}$ is

$$
\begin{aligned}
(4.4) \quad & \mathbb{E}\left(\frac{k'}{n^\alpha} - \frac{k}{n^\alpha} \Big| k\right) \\
&= \frac{1}{n^\alpha}\left[\left(1 - \frac{\alpha k}{n}\right)\left(\frac{\alpha k}{n} + \frac{\beta/s_n}{n}\right) - \frac{\alpha k}{n}\left(\frac{1}{n} - \frac{\beta/s_n}{n^2}\right)(n - \alpha k)\right] \\
&\quad + o\left(\frac{1}{n^{1+\alpha}}\right) \\
&= \frac{\beta/s_n}{n^{1+\alpha}} + o\left(\frac{1}{n^{1+\alpha}}\right).
\end{aligned}
$$

Similarly, the conditional second moment of the increment is

$$
\begin{aligned}
(4.5) \quad & \mathbb{E}\left[\left(\frac{k'}{n^\alpha} - \frac{k}{n^\alpha}\right)^2 \Big| k\right] \\
&= \frac{1}{n^{2\alpha}}\left[\left(1 - \frac{\alpha k}{n}\right)\left(\frac{\alpha k}{n} + \frac{\beta/s_n}{n}\right) + \frac{\alpha k}{n}\left(\frac{1}{n} - \frac{\beta/s_n}{n^2}\right)(n - \alpha k)\right] \\
&\quad + o\left(\frac{1}{n^{1+2\alpha}}\right) \\
&= \frac{2\alpha k}{n^{1+2\alpha}} + o\left(\frac{1}{n^{1+2\alpha}}\right).
\end{aligned}
$$

Since $k \approx sn^{\alpha}$, and recalling that $s_n \to s$ almost surely, we have

$$n^{1+\alpha} \mathbb{E}\left(\frac{k'}{n^{\alpha}} - \frac{k}{n^{\alpha}} \Big| k \right) \to \beta/s$$

and

$$n^{1+\alpha} \mathbb{E}\left[\left(\frac{k'}{n^{\alpha}} - \frac{k}{n^{\alpha}} \right)^2 \Big| k \right] \to 2\alpha s.$$

It is easy to check that all conditional mth moments of $\Delta k/n^{\alpha}$ converge to zero for $m \geq 3$, whence it follows by standard theory [cf., e.g., Karlin and Taylor (1981)] that, as $n \to \infty$, the process $\tilde{K}_n(t) = K_n(\lfloor n^{3/2}t \rfloor)/n^{\alpha}$ converges in distribution to a diffusion process S_t on $[0, \infty)$ with drift β/S_t and diffusion coefficient $\sqrt{2\alpha S_t}$. $\square$

As anticipated, the second order approximation of $g_0(n, k)$ is crucial for establishing the drift of the limiting diffusion, as the first order terms cancel. It is interesting to note that when $\beta = 0$, which yields the normalized stable case, the limiting diffusion reduces to the diffusion approximation of a critical Galton–Watson branching process, also known as the zero-drift Feller diffusion. See, for example, Ethier and Kurtz (1986), Theorem 9.1.3. This also holds approximately for high values of S_t, in which case the drift becomes negligible.

In order to have some heuristics on the behavior of the α-diversity process, Figure 1 shows 3×10^5 steps of the random walk $\{K_n(m)/n^{\alpha}, m \in \mathbb{N}_0\}$ with dynamics as in Theorem 4.1, starting from $1/\sqrt{n}$ with $n = 200$. The three paths correspond to β being equal to 0, 100 and 1000. It is apparent how β influences the dynamic clustering structure in the population.

It is well known that when $\beta = 0$, the point 0 is an absorbing boundary for S_t. The next result provides the boundary classification, using Feller's terminology, for the case $\beta > 0$.

PROPOSITION 4.1. *Let S_t be as in Theorem 4.1 with $\beta > 0$. Then the points 0 and ∞ are, respectively, an entrance and a natural boundary.*

PROOF. The scale function for the process, defined as

$$(4.6) \qquad S(x) = \int_{x_0}^{x} s(y)\, \mathrm{d}y, \qquad 0 < x < \infty,$$

where

$$s(y) = \exp\left\{ -\int_{y_0}^{y} \frac{2\mu(t)}{\sigma^2(t)}\, \mathrm{d}t \right\}$$
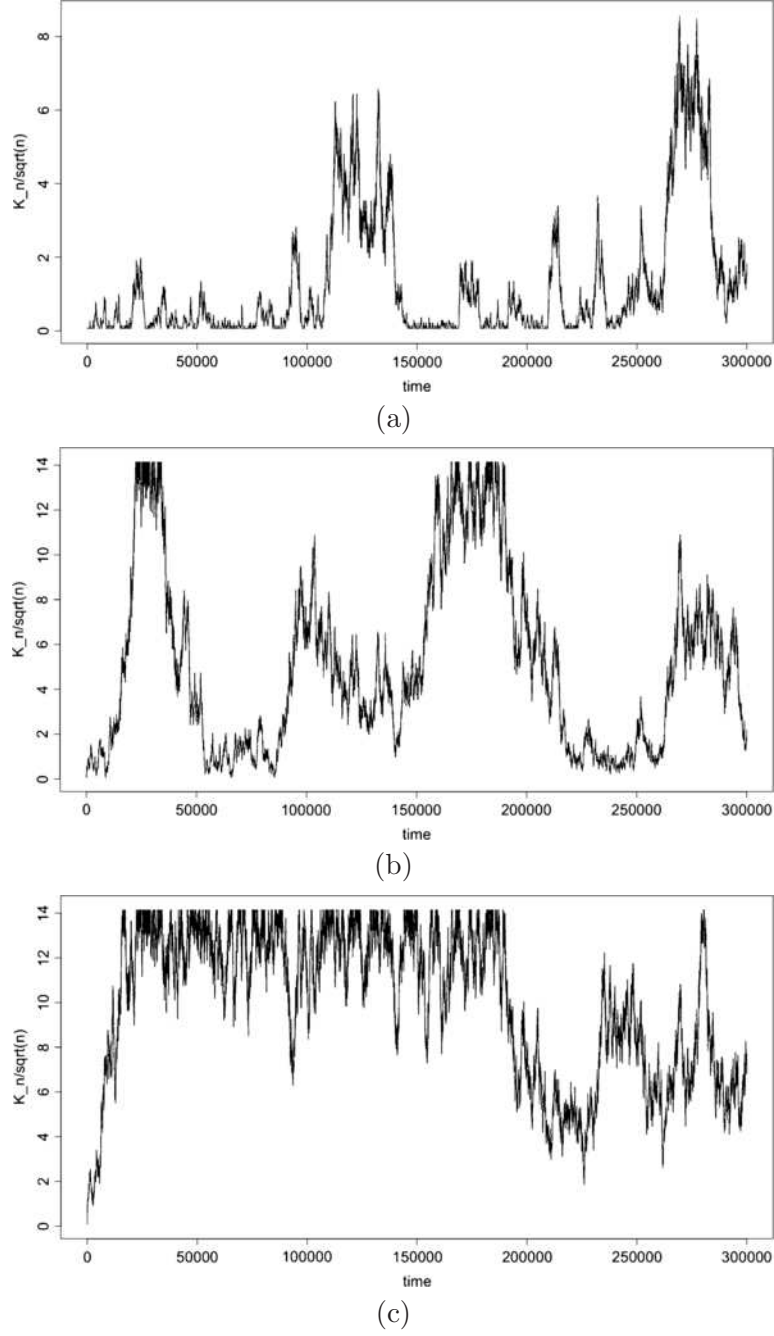
(a)

(b)

(c)

FIG. 1. *Three sample paths of the random walk* $\{K_n(m)/n^\alpha, m \in \mathbb{N}_0\}$, *with dynamics as in Theorem 4.1, starting from* $1/\sqrt{n}$ *with* $n = 200$, *for parameter values:* (a) $\beta = 0$, (b) $\beta = 100$, (c) $\beta = 1000$. *The figures show how* β *influences the dynamic clustering structure in the population.*

and $\mu(x)$ and $\sigma^2(x)$ denote drift and diffusion, equals

$$S(x) = \int_{x_0}^x \exp\left\{-2\beta\left(\frac{1}{y_0} - \frac{1}{y}\right)\right\} \mathrm{d}y$$
$$= e^{-2\beta/y_0}[xe^{2\beta/x} - x_0 e^{2\beta/x_0} - 2\beta\,\mathrm{Ei}(2\beta/x) + \mathrm{Ei}(2\beta/x_0)],$$

where $\mathrm{Ei}(z)$ is the exponential integral

$$\mathrm{Ei}(z) = -\int_{-z}^\infty t^{-1} e^{-t}\,\mathrm{d}t.$$

Letting $S[a,b] = S(b) - S(a)$, for $0 < a < b < \infty$, we have

$$S(0,b] = \lim_{a\downarrow 0} S[a,b] = \infty,$$

(4.7)

$$S[a,\infty) = \lim_{b\uparrow\infty} S[a,b] = \infty.$$

Moreover, the speed measure is given by

$$M[c,d] = \int_c^d [\sigma^2(t)s(t)]^{-1}\,\mathrm{d}t$$
$$= e^{2\beta/y_0}[\mathrm{Ei}(-2\beta/c) - \mathrm{Ei}(-2\beta/d)]$$

from which $M(0,d] = \lim_{c\downarrow 0} M[c,d] < \infty$ and

(4.8) $$M[c,\infty) = \lim_{d\uparrow\infty} M[c,d] = \infty.$$

Now (4.7) implies that

$$\Sigma(0) = \lim_{l\downarrow 0} \int_l^x S(l,y]\,\mathrm{d}M(y) = \infty,$$

$$\Sigma(\infty) = \lim_{l\uparrow\infty} \int_x^r S[y,r)\,\mathrm{d}M(y) = \infty$$

and (4.8) implies

$$N(\infty) = \lim_{r\uparrow\infty} \int_x^r S[x,y]\,\mathrm{d}M(y) = \infty,$$

while

$$N(0) = \lim_{l\downarrow 0} \int_l^x S[y,x]\,\mathrm{d}M(y)$$
$$= \lim_{l\downarrow 0} \int_l^x \frac{e^{-2\beta/y}}{y}\left[xe^{2\beta/x} - ye^{2\beta/y} + 2\beta\left(\mathrm{Ei}\left(\frac{2\beta}{y}\right) - \mathrm{Ei}\left(\frac{2\beta}{x}\right)\right)\right]\mathrm{d}y < \infty$$

since

$$\lim_{y\downarrow 0} \frac{e^{-2\beta/y}}{y} \operatorname{Ei}\left(\frac{2\beta}{y}\right) < \infty.$$

The statement now follows from, for example, Karlin and Taylor (1981), Section 15.6. $\square$

Hence, when $\beta > 0$ neither boundary point is attainable from the interior of the state space, from which the actual state space is $[0, \infty)$ for $\{S_t, t \geq 0\}$ and $(0, \infty)$ for $\{S_t, t > 0\}$. The process can be made to start at 0, in which case it instantly moves toward the interior of the state space and never comes back. Consequently, we will use $(0, \infty)$ or $[0, \infty)$ as state space at convenience, with the agreement that $(0, \infty)$ is referred to $\{S_t, t > 0\}$.

As a corollary, we formalize the well-definedness of the α-diversity diffusion. Denote by $C_0(K)$ the space of continuous functions vanishing at infinity on a locally compact set K, and let $\|\cdot\|$ be a norm which makes $C_0(K)$ a Banach space. Recall that a Feller semigroup on $C_0(K)$ is a one-parameter family of bounded linear operators $\{T(t), t \geq 0\}$ on $C_0(K)$ such that $T(t)$ has the semigroup property $T(s + t) = T(s)T(t)$ for all $s, t \geq 0$, is strongly continuous, that is,

$$\|T(t)f - f\| \to 0 \qquad \text{as } t \to 0, f \in C_0(K),$$

and, for all $t \geq 0$, $T(t)$ is a contraction, that is, $\|T(t)\| \leq 1$, is conservative in the sense that $T(t)1 = 1$, and is positive in the sense that it preserves the cone of nonnegative functions.

COROLLARY 4.1. *For $\beta \geq 0$, let $\mathcal{A}_0$ be the second order differential operator*

$$(4.9) \qquad \mathcal{A}_0 = \frac{\beta}{s}\frac{\mathrm{d}}{\mathrm{d}s} + \frac{1}{2}s\frac{\mathrm{d}^2}{\mathrm{d}s^2}$$

and define

$$(4.10) \quad \mathscr{D}(\mathcal{A}_0) = \{f \in C_0([0, \infty)) \cap C^2((0, \infty)) : \mathcal{A}_0 f \in C_0([0, \infty))\}.$$

Then $\{(f, \mathcal{A}_0 f) : f \in \mathscr{D}(\mathcal{A}_0)\}$ generates a Feller semigroup on $C_0([0, \infty))$.

PROOF. The result follows from Proposition 4.1 together with Corollary 8.1.2 in Ethier and Kurtz (1986). $\square$

An immediate question that arises is whether the α-diversity diffusion is stationary. The following proposition, which concludes the section, provides a negative answer.

PROPOSITION 4.2. *Let $\{S_t, t \geq 0\}$ be as in Theorem 4.1. Then there exists no stationary density for the process.*

PROOF. A stationary density, if it exists, is given by

$$\psi(x) = m(x)[C_1 S(x) + C_2], \qquad x \geq 0,$$

where $m(x) = [s(x)\sigma^2(x)]^{-1}$, $s(x)$ and $S(x)$ are as in (4.6), and C_1, C_2 are constants determined in order to guarantee the nonnegativity and integrability to one of ψ. Here $s(x) = e^{-2\beta/x}$ and

$$S(x) = xe^{2\beta/x} - 2\beta \operatorname{Ei}(2\beta/x)$$

so that

$$\psi(x) = C_1 - 2\beta C_1 x^{-1} e^{-2\beta/x} \operatorname{Ei}(2\beta/x) + C_2 x^{-1} e^{-2\beta/x}.$$

The second term is not integrable in a neighborhood of infinity, since there exists an $x_0 > 0$ such that

$$-x^{-1} e^{-2\beta/x} \operatorname{Ei}(2\beta/x) > x^{-1} \qquad \text{for all } x > x_0,$$

hence C_1 must be zero. Since neither the third term is integrable, this gives the result. $\square$

5. Normalized inverse-Gaussian diffusions. The α-diversity process constructed in the previous section is a key component in the definition of the class of normalized inverse-Gaussian diffusions. In this section we characterize such infinite-dimensional processes in terms of their infinitesimal generator, and show that they can be obtained as the limit in distribution of a certain sequence of Feller diffusions with finitely-many types. The association of the limit family with the class of normalized inverse-Gaussian random probability measures will instead be shown in Section 6.

Consider the $(n-1)$-dimensional simplex

$$\Delta_n = \left\{ z \in [0,1]^n : z_i \geq 0, \sum_{i=1}^n z_i = 1 \right\}$$

and the closed subspace of Δ_n given by

$$\tilde{\Delta}_n = \left\{ z \in [0,1]^n : z_i \geq \varepsilon_n, \sum_{i=1}^n z_i = 1 \right\},$$

so that $\varepsilon_n \leq z_i \leq 1 - (n-1)\varepsilon_n$ for $z_i \in \tilde{\Delta}_n$, where $\{\varepsilon_n\} \subset \mathbb{R}_+$ is a nonincreasing sequence such that

$$(5.1) \qquad 0 < \varepsilon_n < \frac{1}{n} \qquad \forall n \geq 2, \qquad n\varepsilon_n \downarrow 0.$$

Define, for $(z_0, z_1, \ldots, z_n) \in (0, \infty) \times \tilde{\Delta}_n$, the differential operator

$$\mathcal{A}_n = \frac{1}{2} \sum_{i,j=0}^n a_{ij}^{(n)}(z) \frac{\partial^2}{\partial z_i \, \partial z_j} + \frac{1}{2} \sum_{i=0}^n b_i^{(n)}(z) \frac{\partial}{\partial z_i},$$

where the covariance components $(a_{ij}^{(n)}(z))_{i,j=0,\ldots,n}$ are set to be

$$a_{ij}^{(n)}(z) = \begin{cases} z_0, & i=j=0, \\ (z_i - \varepsilon_n)(\delta_{ij}(1 - n\varepsilon_n) - (z_j - \varepsilon_n)), & 1 \le i,j \le n, \\ 0, & \text{else}, \end{cases}$$

and, for $\beta > 0$, the drift components are

$$b_0^{(n)}(z) = \frac{\beta}{z_0},$$

$$b_i^{(n)}(z) = \frac{\beta}{z_0(n-1)}(1 - z_i) - \frac{\beta}{z_0}z_i$$
$$- \alpha(1 - \exp\{-(z_i - \varepsilon_n)e^{1/\varepsilon_n}\}), \qquad i = 1, \ldots, n.$$

Observe that $a_{ij}^{(n)}(z)$, for $1 \le i, j \le n$, can be seen as a Wright–Fisher type covariance restricted to $[\varepsilon_n, 1 - (n-1)\varepsilon_n]^n$, since for such indices i, j

$$(5.2) \qquad a_{ij}^{(n)}(z) = \begin{cases} (z_i - \varepsilon_n)(1 - (n-1)\varepsilon_n - z_i), & i = j, \\ -(z_i - \varepsilon_n)(z_j - \varepsilon_n), & i \ne j, \end{cases}$$

and that the first two terms in $b_i^{(n)}(z)$, $i = 1, \ldots, n$, equal

$$\frac{\beta}{z_0(n-1)}(1 - (n-1)\varepsilon_n - z_i) - \frac{\beta}{z_0}(z_i - \varepsilon_n)$$

from which the behavior at the boundary is clear. For ease of exposition and in analogy with the previous section, whenever convenient we will denote z_0 by s, so that for instance $\mathcal{A}_n$ can be written more explicitly,

$$(5.3) \quad \mathcal{A}_n = \frac{\beta}{s}\frac{\partial}{\partial s} + \frac{1}{2}s\frac{\partial^2}{\partial s^2} + \frac{1}{2}\sum_{i,j=1}^n a_{ij}^{(n)}(z)\frac{\partial^2}{\partial z_i\,\partial z_j} + \frac{1}{2}\sum_{i=1}^n b_i^{(n)}(z)\frac{\partial}{\partial z_i}$$

with $a_{ij}^{(n)}(z)$ and $b_i^{(n)}(z)$ as above. The domain of $\mathcal{A}_n$ is taken to be

$$(5.4) \qquad \mathscr{D}(\mathcal{A}_n) = \{f : f = f_0 \times f_1, f_0 \in \mathscr{D}(\mathcal{A}_0), f_1 \in C^2(\tilde{\Delta}_n)\},$$

where $(f_0 \times f_1)(s, z) = f_0(s)f_1(z)$, $\mathscr{D}(\mathcal{A}_0)$ is (4.10), and

$$C^2(\tilde{\Delta}_n) = \{f \in C(\tilde{\Delta}_n) : \exists \tilde{f} \in C^2(\mathbb{R}^n), \tilde{f}|_{\tilde{\Delta}_n} = f\}.$$

The operator $\mathcal{A}_n$ drives $n+1$ components: those labeled from 1 to n can be seen as the frequencies associated to n species in a large population, bounded from below by ε_n; the z_0 or s component is a positive real variable which evolves independently according to the α-diversity diffusion (4.2) and contributes to drive the drift of the other n components.

Denote by $C_0([0,\infty) \times \tilde{\Delta}_n)$ the Banach space of continuous functions on $[0,\infty) \times \tilde{\Delta}_n$ which vanish at infinity, equipped with the supremum norm

$\|f\| = \sup_{x \in [0,\infty) \times \tilde{\Delta}_n} f(x)$, and by $\mathscr{P}(B)$ the set of Borel probability measures on B. Recall that a Markov process $\{X(t), t \geq 0\}$, taking values in a metric space E, is said to correspond to a semigroup $\{T(t)\}$, acting on a closed subspace L of the space of bounded functions on E, if

$$\mathbb{E}[f(X(t+s))|\mathscr{F}_t^X] = T(s)f(X(t)), \qquad s, t \geq 0,$$

for every $f \in L$, where $\mathscr{F}_t^X = \sigma(X(u), u \leq t)$.

PROPOSITION 5.1. *Let $\mathcal{A}_n$ be the operator defined in (5.3) and (5.4). The closure in $C_0([0,\infty) \times \tilde{\Delta}_n)$ of $\mathcal{A}_n$ generates a strongly continuous, positive, conservative, contraction semigroup $\{\mathcal{S}_n(t)\}$ on $C_0([0,\infty) \times \tilde{\Delta}_n)$. For every $\nu_n \in \mathscr{P}([0,\infty) \times \tilde{\Delta}_n)$ there exists a strong Markov process $Z^{(n)}(\cdot) = \{Z^{(n)}(t), t \geq 0\}$ corresponding to $\{\mathcal{S}_n(t)\}$ with initial distribution ν_n and sample paths in $C_{[0,\infty) \times \tilde{\Delta}_n}([0,\infty))$ with probability one.*

PROOF. We proceed by verifying the hypothesis of the Hille–Yosida theorem. Note first that $\mathcal{A}_n$ satisfies the positive maximum principle, that is, for $f \in \mathcal{D}(\mathcal{A}_n)$ and $(s_*, z_*) \in (0, \infty) \times \tilde{\Delta}_n$ such that $\|f\| = f(s_*, z_*) \geq 0$ we have $\mathcal{A}_n f(s_*, z_*) \leq 0$. Indeed, writing $\mathcal{A}_n = \mathcal{A}_0 + \mathcal{A}_{n,1}$ to indicate the first two and last two terms in (5.3), it is immediate to check that $\mathcal{A}_0$ and $\mathcal{A}_{n,1}$ satisfy the positive maximum principle on $[0, \infty)$ and $\tilde{\Delta}_n$, respectively. If $f_0(s_*) \geq 0, f_1(z_*) \geq 0$, then $\mathcal{A}_0 f_0(s) \leq 0$ and $\mathcal{A}_{n,1} f_1(z) \leq 0$, while if $f_0(s_*) \leq 0, f_1(z_*) \leq 0$, then $\mathcal{A}_0 f_0(s_*) \geq 0$ and $\mathcal{A}_{n,1} f_1(z_*) \geq 0$. In both cases

$$\mathcal{A}_n f(s_*, z_*) = f_1(z_*) \mathcal{A}_0 f_0(s_*) + f_0(s_*) \mathcal{A}_{n,1} f_1(z_*) \leq 0.$$

Let now $L \subset \mathcal{D}(\mathcal{A}_n)$ be the algebra generated by functions $f = f_0 \times f_1$, with $f_0 \in \mathcal{D}(\mathcal{A}_0)$, $\mathcal{D}(\mathcal{A}_0)$ as in (4.10) and $f_1 = z^c = z_1^{c_1} \cdots z_n^{c_n} \in C^2(\tilde{\Delta}_n)$, $c_i \in \mathbb{N}_0$, so that L is dense in $C_0([0,\infty) \times \tilde{\Delta}_n)$, and so is $\mathcal{D}(\mathcal{A}_n)$. Denoting $c + d_{(i)} = (c_0, \ldots, c_i + d, \ldots, c_n)$, for $f \in L$ we have

$$
\begin{aligned}
\mathcal{A}_n(&f_0(s) \times z^c) \\
&= f_1(z) \mathcal{A}_0 f(s) \\
&\quad + \frac{f_0(z_0)}{2} \Bigg\{ \sum_{i=1}^n \Bigg[c_i(c_i - 1)(-z^c + (1 - (n-2)\varepsilon_n)\varepsilon_n z^{c-1_{(i)}}) \\
&\qquad\qquad\qquad\qquad - \varepsilon_n(1 - (n-1)\varepsilon_n) z^{c-2_{(i)}}) \\
&\quad + \sum_{j \neq i}^n c_i c_j (-z^c + \varepsilon_n(z^{c-1_{(j)}} + z^{c-1_{(i)}}) \\
&\qquad\qquad\qquad\qquad - \varepsilon_n z^{c-1_{(i)} - 1_{(j)}})) \Bigg]
\end{aligned}
$$

(5.5)

$$+ \sum_{i=1}^{n} c_i \left[\frac{\beta}{z_0(n-1)} (z^{c-1}(i) - nz^c) \right.$$

$$\left. - \alpha z^{c-1}(i) + \alpha e^{\varepsilon_n \exp\{1/\varepsilon_n\}} e^{-\exp\{1/\varepsilon_n\} z_i} z^{c-\delta_i} \right] \Bigg\},$$

so that the image of $\mathcal{A}_n$ contains functions of type $f_0 \times z^c$ and $f_0 \times e^{-b_0 z_i} z^c$, with b_0 fixed. For every $g(x) \in C(K)$, with K compact, and $f(x) = e^{b_0 x} g(x) \in C(K)$, there exists a sequence $\{p^{(k)}\}$ of polynomials on K such that $\|f - p^{(k)}\| \to 0$, so that $\|e^{-b_0 z} p^{(k)} - g\| \to 0$. It follows that the image of $\mathcal{A}_n$ is dense in $C_0([0,\infty) \times \tilde{\Delta}_n)$, and so is that of $\lambda - \mathcal{A}_n$ for all but at most countably many $\lambda > 0$. The first assertion now follows from Theorem 4.2.2 of Ethier and Kurtz (1986) and by noting that $1 \in \mathscr{D}(\mathcal{A}_n)$ and $\mathcal{A}_n 1 = 0$, that is, $\mathcal{A}_n$ is conservative. The second assertion with $D_{[0,\infty) \times \tilde{\Delta}_n}([0,\infty))$, the space of right-continuous functions with left limits, in place of $C_{[0,\infty) \times \tilde{\Delta}_n}([0,\infty))$, follows from Theorem 4.2.7 of Ethier and Kurtz (1986). To prove the almost sure continuity of sample paths, it is enough to show that for every $z^* \in (0,\infty) \times \tilde{\Delta}_n$ and $\epsilon > 0$ there exists a function $f \in \mathscr{D}(\overline{\mathcal{A}}_n)$ such that

$$f(z^*) = \|f\|, \qquad \sup_{z \in B(z^*,\epsilon)^c} f(z) < f(z^*), \qquad \overline{\mathcal{A}}_n f(z^*) = 0,$$

where $B(z^*,\epsilon)^c$ is the complement of an ϵ-neighborhood of z^* in the topology of coordinatewise convergence [cf. Ethier and Kurtz (1986), Remark 4.2.10]. This can be done by means of a function $f \in \mathscr{D}(\mathcal{A}_n)$ which is flat in z^* and rapidly decreasing away from z^*, for example, of type $f(z) = c_1 - c_2 \sum_{i=0}^{n} (z_i - z_i^*)^4$ for appropriate constants c_1, c_2. $\square$

For $Z^{(n)}(\cdot)$ as in Proposition 5.1, consider now the mapping $\rho_n(Z^{(n)}(\cdot))$, where $\rho_n : [0,\infty) \times \tilde{\Delta}_n \to [0,\infty) \times \overline{\nabla}_\infty$ is defined as

$$(5.6) \qquad \rho_n(z) = (z_0, z_{(1)}, z_{(2)}, \ldots, z_{(n)}, 0, 0, \ldots),$$

$(z_{(1)}, \ldots, z_{(n)})$ is the vector of decreasingly ordered statistics of $(z_1, \ldots, z_n) \in \tilde{\Delta}_n$, and $\overline{\nabla}_\infty$ is the closure of the infinite-dimensional ordered simplex, defined in (1.1). The following proposition states that $\rho_n(Z^{(n)}(\cdot))$ is still a well-defined Markov process. Define

$$\tilde{\nabla}_n = \left\{ z \in \overline{\nabla}_\infty : z_n \geq \varepsilon_n > z_{n+1} = 0, \sum_{i=1}^{n} z_i = 1 \right\}$$

and observe that $z \in \overline{\nabla}_\infty$ satisfies $z_i \leq 1/i$ for all i, so that $\tilde{\nabla}_n$ is nonempty by (5.1).

PROPOSITION 5.2. *Let $\tilde{\mathcal{A}}_n$ be defined by the right-hand side of (5.3), with domain*

$$\mathscr{D}(\tilde{\mathcal{A}}_n) = \{f : f = f_0 \times f_1, f_0 \in \mathscr{D}(\mathcal{A}_0), f_1 \in C^2_{\rho_n}(\tilde{\nabla}_n)\},$$

where $\mathscr{D}(\mathcal{A}_0)$ is as in (4.10) and

$$C^2_{\rho_n}(\tilde{\nabla}_n) = \{f \in C(\tilde{\nabla}_n) : f \circ \rho_n \in C^2(\tilde{\Delta}_n)\}.$$

Then the closure of $\tilde{\mathcal{A}}_n$ in $C_0([0, \infty) \times \tilde{\nabla}_n)$ generates a strongly continuous, positive, conservative, contraction semigroup $\{\mathcal{T}_n(t)\}$ on $C_0([0, \infty) \times \tilde{\nabla}_n)$. For every $\nu_n \in \mathscr{P}([0, \infty) \times \tilde{\Delta}_n)$, let $Z^{(n)}$ be as in Proposition 5.1. Then $\rho_n(Z^{(n)}(\cdot))$ is a strong Markov process corresponding to $\{\mathcal{T}_n(t)\}$ with initial distribution $\nu_n \circ \rho_n^{-1}$ and sample paths in $C_{[0,\infty) \times \tilde{\nabla}_n}([0, \infty))$ with probability one.

PROOF. The result follows from Proposition 2.4 in Ethier and Kurtz (1981), with Δ_n, ∇_n and ρ_n there substituted by $[0, \infty) \times \tilde{\Delta}_n$, $[0, \infty) \times \tilde{\nabla}_n$ and (5.6), respectively. $\square$

We now turn the attention to the limit of $\rho_n(Z^{(n)}(\cdot))$ when the number of types goes to infinity. To this end, consider that $\overline{\nabla}_\infty$ is a compact and metrizable space in the topology of coordinatewise convergence, and let $C_0([0, \infty) \times \overline{\nabla}_\infty)$ be the Banach space of continuous functions on $[0, \infty) \times \overline{\nabla}_\infty$ which vanish at infinity, with the supremum norm $\|f\| = \sup_{z \in [0,\infty) \times \overline{\nabla}_\infty} |f(z)|$. The key issue for showing that the closure of the differential operator $\mathcal{A}$, defined in (1.4), generates a Feller diffusion on $C_0([0, \infty) \times \overline{\nabla}_\infty)$ is the choice of the domain of $\mathcal{A}$. Here we adapt to the present framework a technique indicated by Ethier and Kurtz (1981). Consider polynomials $\varphi_m : \overline{\nabla}_\infty \to [0, 1]$ defined as

$$(5.7) \qquad \varphi_1(z) = 1, \qquad \varphi_m(z) = \sum_{i=1}^{\infty} z_i^m, \qquad m \geq 2.$$

Since $z \in \overline{\nabla}_\infty$ implies $z_i \leq i^{-1}$, functions φ_m with $m \geq 2$ are uniformly convergent, and sums in (1.4) are assumed to be computed on

$$(5.8) \qquad \nabla_\infty = \left\{ z = (z_1, z_2, \ldots) : z_1 \geq z_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} z_i = 1 \right\}$$

and extended to $\overline{\nabla}_\infty$ by continuity, so that, for example,

$$\sum_{i=1}^{\infty} (1 + z_i) \frac{\partial}{\partial z_i} \varphi_2(z) = 2 + 2\varphi_2(z)$$

instead of $2 \sum_{i=1}^{\infty} z_i + 2\varphi_2(z)$. Write

$$(5.9) \qquad \mathcal{A} = \mathcal{A}_0 + \mathcal{A}_1$$

to indicate the first two and last two terms in (1.4), and denote

$$(5.10) \quad \mathscr{D}(\mathcal{A}_1) = \{\text{sub-algebra of } C(\overline{\nabla}_\infty) \text{ generated by } \varphi_m \text{ as in (5.7)}\}.$$

The domain $\mathscr{D}(\mathcal{A})$ of the operator (1.4) is then taken to be

(5.11)
$$\mathscr{D}(\mathcal{A}) = \{\text{sub-algebra of } C_0([0,\infty) \times \overline{\nabla}_\infty) \text{ generated by }$$
$$f = f_0 \times f_1 : f_0 \in \mathscr{D}(\mathcal{A}_0), f_1 \in \mathscr{D}(\mathcal{A}_1)\}$$

with $\mathscr{D}(\mathcal{A}_0)$ as in (4.10) and $\mathscr{D}(\mathcal{A}_1)$ as above.

LEMMA 5.1. *The sub-algebra $\mathscr{D}(\mathcal{A}_1) \subset C(\overline{\nabla}_\infty)$ is dense in $C(\overline{\nabla}_\infty)$.*

PROOF. See the proof of Theorem 2.5 in Ethier and Kurtz (1981). □

We also need the following lemma, which shows that the operator $\mathcal{A}_1$ is triangulizable.

LEMMA 5.2. *Let $\mathcal{A}_1$ be as in (5.9) and, for any $m \geq 2$, let L_m be the algebra generated by polynomials as in (5.7), with degree not greater than m. Then $\mathcal{A}_1 : L_m \to L_m$.*

PROOF. The assertion follows from equation (2.4) in Feng and Sun (2010), with θ replaced by β/s. □

Then we have the following result.

THEOREM 5.1. *Let $\mathcal{A}$ be the operator defined by (1.4) and (5.11). The closure in $C_0([0,\infty) \times \overline{\nabla}_\infty)$ of $\mathcal{A}$ generates a strongly continuous, positive, conservative, contraction semigroup $\{\mathcal{T}(t)\}$ on $C_0([0,\infty) \times \overline{\nabla}_\infty)$. For every $\nu \in \mathscr{P}([0,\infty) \times \overline{\nabla}_\infty)$, there exists a strong Markov process $Z(\cdot)$ corresponding to $\{\mathcal{T}(t)\}$ with initial distribution ν and sample paths in $C_{[0,\infty) \times \overline{\nabla}_\infty}([0,\infty))$ with probability one.*

PROOF. For every $g \in C_0([0,\infty) \times \overline{\nabla}_\infty)$, define $r_n : C_0([0,\infty) \times \overline{\nabla}_\infty) \to C_0([0,\infty) \times \tilde{\nabla}_n)$ to be the bounded linear map

$$r_n g = g|_{[0,\infty) \times \tilde{\nabla}_n}$$

given by the restriction of g to $[0,\infty) \times \tilde{\nabla}_n$. Note that $r_n : \mathscr{D}(\mathcal{A}) \to \mathscr{D}(\tilde{\mathcal{A}}_n)$, with $\tilde{\mathcal{A}}_n$ as in Proposition 5.2, and that

(5.12)
$$\|r_n g - g\| \longrightarrow 0, \qquad g \in C_0([0,\infty) \times \overline{\nabla}_\infty).$$

Then, for $g \in \mathscr{D}(\mathcal{A})$ and $z \in (0,\infty) \times \tilde{\nabla}_n$, we have

$$|\tilde{\mathcal{A}}_n r_n g(z) - r_n \mathcal{A} g(z)|$$
$$= \frac{1}{2} \left| \sum_{i,j=1}^n (a_{ij}^{(n)}(z) - z_i(\delta_{ij} - z_j)) \frac{\partial^2 g(z)}{\partial z_i \, \partial z_j} \right.$$

$$+ \sum_{i=1}^{n} \left(\frac{\beta}{z_0(n-1)} (1 - z_i) - \alpha(\exp\{-(z_i - \varepsilon_n)e^{1/\varepsilon_n}\}) \right) \frac{\partial g(z)}{\partial z_i} \Bigg|$$

with $a_{ij}^{(n)}(z)$ as in (5.2). In particular,

$$|a_{ij}^{(n)}(z) - z_i(\delta_{ij} - z_j)|$$

$$= \begin{cases} \varepsilon_n[(z_i - \varepsilon_n)(n-1) + 1 - z_i], & i = j, \\ \varepsilon_n[z_i + z_j - \varepsilon_n], & i \neq j, \end{cases}$$

which is bounded above by $n\varepsilon_n$, from which

$$|\tilde{\mathcal{A}}_n r_n g(z) - r_n \mathcal{A} g(z)|$$

(5.13)
$$\leq n\varepsilon_n \sum_{i,j=1}^{n} \left| \frac{\partial^2 g(z)}{\partial z_i \partial z_j} \right| + \frac{\beta}{z_0(n-1)} \sum_{i=1}^{n} \left| \frac{\partial g(z)}{\partial z_i} \right|$$

$$+ \sum_{i=1}^{n} \exp\{-(z_i - \varepsilon_n)e^{1/\varepsilon_n}\} \left| \frac{\partial g(z)}{\partial z_i} \right|.$$

For $g \in \mathcal{D}(\mathcal{A})$ of type $g = f_0 \times f_1$, with $f_0 \in \mathcal{D}(\mathcal{A}_0)$ and $f_1 = \varphi_{m_1} \cdots \varphi_{m_k}$, we have

(5.14)
$$\sum_{i=1}^{n} \left| \frac{\partial g(z)}{\partial z_i} \right| = |f_0(z_0)| \sum_{i=1}^{n} \sum_{j=1}^{k} m_j z_i^{m_j - 1} \prod_{h \neq j} \varphi_{m_h}$$

$$\leq |f_0(z_0)| \sum_{j=1}^{k} m_j \sum_{i=1}^{n} z_i^{m_j - 1}$$

so that

$$\sum_{i=1}^{n} \exp\{-(z_i - \varepsilon_n)e^{1/\varepsilon_n}\} \left| \frac{\partial g(z)}{\partial z_i} \right| \leq n\varepsilon_n |f_0(z_0)| \sum_{j=1}^{k} m_j \to 0$$

uniformly as $n \to \infty$ by (5.1). Furthermore,

$$\sum_{i,j=1}^{n} \left| \frac{\partial^2 g(z)}{\partial z_i \partial z_j} \right|$$

$$\leq |f_0(z_0)| \sum_{i,j=1}^{\infty} \left[\partial_{ij} \varphi_{m_h} \prod_{\ell \neq h} \varphi_{m_\ell} + \sum_{q \neq h} \partial_i \varphi_{m_h} \partial_j \varphi_{m_q} \prod_{\ell \neq h, q} \varphi_{m_\ell} \right]$$

$$= |f_0(z_0)| \left[m_h(m_h - 1) \varphi_{m_h - 2} \prod_{\ell \neq h} \varphi_{m_\ell} \right.$$

$$+ \sum_{q \neq h} m_h m_q \varphi_{m_h + m_q - 2} \prod_{\ell \neq h, q} \varphi_{m_\ell}$$

$$\left. + \sum_{q \neq h} m_h m_q \varphi_{m_h - 1} \varphi_{m_q - 1} \prod_{\ell \neq h, q} \varphi_{m_\ell} \right]$$

$$\leq |f_0(z_0)| \left[m_h(m_h - 1) + 2 \sum_{q \neq h} m_h m_q \right],$$

whose right-hand side is bounded. Since also the right-hand side of (5.14) is bounded above by $|f_0(z_0)| \sum_{j=1}^k m_j$, it follows by (5.1) that the right-hand side of (5.13) goes to zero uniformly and, by means of (5.12), that

$$(5.15) \qquad \|\tilde{\mathcal{A}}_n r_n g - \mathcal{A} g\| \longrightarrow 0, \qquad g \in \mathscr{D}(\mathcal{A}).$$

Proposition 5.2 implies that $\tilde{\mathcal{A}}_n$ is a dissipative operator for every $n \geq 1$, so that by (5.15) $\mathcal{A}$ is dissipative. Moreover, Lemma 5.1 and Lemma 5.2, respectively, imply that $\mathscr{D}(\mathcal{A})$ and the range of $\lambda - \mathcal{A}$, for all but at most countably many $\lambda > 0$, are dense in $C_0([0, \infty) \times \overline{\nabla}_\infty)$. The fact that the closure of $\mathcal{A}$ generates a strongly continuous contraction semigroup $\{\mathcal{T}(t)\}$ on $C_0([0, \infty) \times \overline{\nabla}_\infty)$ now follows from the Hille–Yosida theorem [see Theorem 1.2.12 in Ethier and Kurtz (1986)]. It is also immediate to check that $\mathcal{A} 1 = 0$, so that $(1, 0) \in \overline{(g, \mathcal{A} g)}$ and $\{\mathcal{T}(t)\}$ is conservative. Finally, (5.15), together with Lemma 5.1 and Theorem 1.6.1 of Ethier and Kurtz (1986), implies the semigroup convergence

$$(5.16) \qquad \|\mathcal{T}_n(t) r_n g - \mathcal{T}(t) g\| \longrightarrow 0, \qquad g \in C_0([0, \infty) \times \overline{\nabla}_\infty),$$

uniformly on bounded intervals. From Proposition 5.2, $\{\mathcal{T}_n(t)\}$ is a positive operator for every $n \geq 1$, so that $\{\mathcal{T}(t)\}$ is in turn positive.

The second assertion of the theorem, with $D_{[0,\infty) \times \overline{\nabla}_\infty}([0, \infty))$ in place of $C_{[0,\infty) \times \overline{\nabla}_\infty}([0, \infty))$, follows from Theorem 4.2.7 in Ethier and Kurtz (1986), while the continuity of sample paths follows from a similar argument to that used in the proof of Proposition 5.1. $\square$

The following corollary formalizes the convergence in distribution of the sequence of processes of Proposition 5.2 to the infinite-dimensional diffusion of Theorem 5.1.

COROLLARY 5.1. *Let $Z^{(n)}(\cdot)$ be as in Proposition 5.1 with initial distribution $\nu_n \in \mathscr{P}([0, \infty) \times \tilde{\Delta}_n)$, and let $Z(\cdot)$ be as in Theorem 5.1 with initial distribution $\nu \in \mathscr{P}([0, \infty) \times \overline{\nabla}_\infty)$. If $\nu_n \circ \rho_n^{-1} \Rightarrow \nu$ on $\overline{\nabla}_\infty$, then*

$$\rho_n(Z^{(n)}(\cdot)) \Rightarrow Z(\cdot) \qquad in \ C_{[0,\infty) \times \overline{\nabla}_\infty}([0, \infty))$$

as $n \to \infty$.

PROOF. The result with $D_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ in place of $C_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ follows from Proposition 5.2, together with (5.16) and Theorem 4.2.5 in Ethier and Kurtz (1986). The fact that the weak convergence holds in $C_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ follows from relativization of the Skorohod topology. $\square$

6. A population model for normalized inverse-Gaussian diffusions. By formalizing the population process briefly mentioned in Section 4 for constructing the α-diversity diffusion, in this section we provide a discrete approximation, based on a countable number of particles, for the diffusion with operator (1.4). More specifically, this is obtained as the limit in distribution of the process of frequencies of types associated with a set of particles sampled from a normalized inverse-Gaussian random probability measure, jointly with the normalized version of the diversity process.

In view of (2.4), the conditional distribution of the ith component of an exchangeable sequence $(X_1,\ldots,X_n)$ drawn from a random probability measure of Gibbs type can be written

$$
\begin{aligned}
(6.1) \quad & \mathbb{P}\{X_i \in \cdot | X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n\} \\
& = g_0(n-1,K_{n-1,i})\nu_0(\cdot) \\
& \quad + g_1(n-1,K_{n-1,i})\sum_{j=1}^{K_{n-1,i}}(n_j-\alpha)\delta_{X_j^*}(\cdot),
\end{aligned}
$$

where ν_0 is a nonatomic probability measure and $(X_1^*,\ldots,X_{K_{n-1,i}}^*)$ are the $K_{n-1,i}$ distinct values in $(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$. For fixed n, define a Markov chain $\{X^{(n)}(m),m\geq 0\}$ on $\mathbb{X}^n$ by means of the transition semigroup

$$
T_n f(x) = \int f(y)p_n(x,\mathrm{d}y), \qquad f \in C_0(\mathbb{X}^n),
$$

where $x,y \in \mathbb{X}^n$, $C_0(\mathbb{X}^n)$ is the space of Borel-measurable continuous functions on $\mathbb{X}^n$ vanishing at infinity,

$$
(6.2) \qquad p_n(x,\mathrm{d}y) = \frac{1}{n}\sum_{i=1}^n \tilde{p}_1(\mathrm{d}y_i|x_{(-i)})\prod_{k\neq i}\delta_{x_k}(\mathrm{d}y_k),
$$

$x_{(-i)} = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$ and $\tilde{p}_1(\mathrm{d}y|x_{(-i)})$ is (6.1). The interpretation is as follows. At each transition one component is selected at random with uniform probability, and is updated with a value sampled from (6.1), conditional on all other components, which are left unchanged. Hence, the incoming particle is either a new type (a mutant offspring) or a copy of an old type (a copied offspring). Embed now the chain in a pure jump Markov process on $\mathbb{X}^n$ with exponentially distributed waiting times with intensity one, and denote the resulting process by $X^{(n)}(\cdot) = \{X^{(n)}(t),t\geq 0\}$. The in-

finitesimal generator of $X^{(n)}(\cdot)$ is given by

$$B_n f(x) = \frac{1}{n} \sum_{i=1}^{n} g_0^{(n-1,i)} [P_i f(x) - f(x)]$$

(6.3)

$$+ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_{n-1,i}} g_1^{(n-1,i)} (n_j - \alpha) [\Phi_{j*i} f(x) - f(x)]$$

with domain

(6.4) $$\mathscr{D}(B_n) = \{ f : f \in C_0(\mathbb{X}^n) \}.$$

Here $\Phi_{j*i} : C_0(\mathbb{X}^n) \to C_0(\mathbb{X}^{n-1})$ is defined as

(6.5) $$\Phi_{j*i} f(x_1, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, x_j^*, x_{i+1}, \ldots, x_n)$$

for $x_j^* \in (x_1^*, \ldots, x_{K_{n-1,i}}^*)$, P is the transition semigroup

(6.6) $$Pg(z) = \int g(y) p_1(z, \mathrm{d}y), \qquad g \in C_0(\mathbb{X}),$$

where $p_1(z, \mathrm{d}y)$ is given by

(6.7) $$p_1(z, \mathrm{d}y) = \nu_0(\mathrm{d}y),$$

$P_i f$ denotes P acting on the ith coordinate of f, and we have set for brevity

(6.8) $$g_j^{(n-1,i)} = g_j(n-1, K_{n-1,i}), \qquad j = 0, 1.$$

Defining (6.6) and (6.7) separately is somewhat redundant, but will allow us to provide a general expression for the global mutation rate in this particle representation before making the assumptions of nonatomicity and parent independence as in (6.7). See (A.15) below.

Define now the map $w : \mathbb{X}^n \to \overline{\nabla}_\infty$ by

(6.9) $$w(x) = w(x^{(n)}) = (z_1, \ldots, z_{K_n}, 0, 0, \ldots),$$

where z_j and K_n, respectively, denote the relative frequency of the jth most abundant type and the number of types in $X^{(n)}$. Let also $\mathcal{A}$ be as in (1.4). The next theorem states that

(6.10) $$[K_n(\cdot)/n^\alpha, w(X^{(n)}(\cdot))] = \{ [K_n(t)/n^\alpha, w(X^{(n)}(t))], t \geq 0 \},$$

if appropriately rescaled in time, converges in distribution to the process with generator $\overline{\mathcal{A}}$. The proof is deferred to the Appendix and contains, as a byproduct, a more formal derivation of Theorem 4.1.

THEOREM 6.1. *Let $X^{(n)}(\cdot)$ be the $\mathbb{X}^n$-valued process with generator (6.3) and (6.4), $w\colon\mathbb{X}^n \to \overline{\nabla}_\infty$ as in (6.9) and $Z(\cdot)$ as in Theorem 5.1. If*

$$(6.11) \qquad [K_n(0)/n^\alpha, w(X^{(n)}(0))] \Rightarrow Z(0),$$

then

$$(6.12) \qquad [K_n(n^{3/2}t)/n^\alpha, w(X^{(n)}(n^2t/2))] \Rightarrow Z(t)$$

in $C_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$.

We conclude the section by showing the reversibility of the particle process. Denote the joint distribution of an n-sized sequence from the generalized Pólya urn scheme (2.4) by

$$\mathcal{M}_n(\mathrm{d}x_1,\dots,\mathrm{d}x_n)$$
$$= \nu_0(\mathrm{d}x_1)\prod_{i=1}^{n-1}\left[g_0(i,K_i)\nu_0(\mathrm{d}x_{i+1}) + g_1(i,K_i)\sum_{j=1}^{K_i}(n_j-\alpha)\delta_{x_j^*}(\mathrm{d}x_{i+1})\right].$$

PROPOSITION 6.1. *Let $X^{(n)}(\cdot)$ be the $\mathbb{X}^n$-valued process with generator given by (6.3) and (6.4). Then $X^{(n)}(\cdot)$ is reversible with respect to $\mathcal{M}_n$.*

PROOF. Let $q_n(x,\mathrm{d}y)$ denote the infinitesimal transition kernel on $\mathbb{X}^n \times \mathcal{B}(\mathbb{X}^n)$ of $X^{(n)}$. Denoting by λ_n the rate at which the discontinuities of $X^{(n)}$ occur, and recalling (6.2), we have

$$\mathcal{M}_n(\mathrm{d}x)q_n(x,\mathrm{d}y)$$
$$= \mathcal{M}_n(\mathrm{d}x)\lambda_n\frac{1}{n}\sum_{i=1}^{n}p_1(\mathrm{d}y_i|x_{(-i)})\prod_{k\neq i}\delta_{x_k}(\mathrm{d}y_k)$$
$$= \frac{\lambda_n}{n}\sum_{i=1}^{n}\mathcal{M}_{n-1}(\mathrm{d}x_{(-i)})p_1(\mathrm{d}x_i|x_{(-i)})p_1(\mathrm{d}y_i|x_{(-i)})\prod_{k\neq i}\delta_{x_k}(y_k)$$
$$= \frac{\lambda_n}{n}\sum_{i=1}^{n}\mathcal{M}_{n-1}(\mathrm{d}y_{(-i)})p_1(\mathrm{d}x_i|y_{(-i)})p_1(\mathrm{d}y_i|y_{(-i)})\prod_{k\neq i}\delta_{y_k}(x_k)$$
$$= \mathcal{M}_n(\mathrm{d}y)\frac{1}{n}\sum_{i=1}^{n}\lambda_n p_n(\mathrm{d}x_i|y_{-i})\prod_{k\neq i}\delta_{y_k}(x_k) = \mathcal{M}_n(\mathrm{d}y)q_n(y,\mathrm{d}x),$$

giving the result. $\square$

7. Conditioning on the alpha-diversity. We conclude by discussing an interesting connection with the two-parameter model (1.2). In the Introduction it was observed that conditioning on the α-diversity diffusion S_t to be con-

stant in the operator (1.4) only yields the special case $\beta = 0$ and $\alpha = 1/2$, consistently with the associated random probability measures. It turns out that performing the same conditioning operation in the particle construction of the previous section, before taking the limit for $n \to \infty$, yields a particular instance of the two-parameter model. The following proposition states that under this pre-limit conditioning with $S_t \equiv s$, the normalized inverse-Gaussian model with operator (1.4) reduces to the two-parameter model with $(\theta, \alpha) = ((\alpha s)^2, \alpha)$ and $\alpha = 1/2$.

PROPOSITION 7.1. *Let $X^{(n)}(\cdot)$ be as in Theorem 6.1, w be as in (6.9), $\tilde{Z}^{(n)}(\cdot)$ be defined by the left-hand side of (6.12), and denote by $V^{\theta,\alpha}(\cdot)$ the process with operator $\mathcal{L}^{\theta,\alpha}$ as in (1.2). Then*

$$[\tilde{Z}^{(n)}(\cdot)|S_t \equiv s] \Rightarrow V^{s^2/4,1/2}(\cdot)$$

in $C_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ as $n \to \infty$.

PROOF. Let $\alpha = 1/2$ throughout the proof. In the pre-limit version of the process of frequencies derived from the particle process, that is, (6.9), conditioning on $S_t \equiv s$ means conditioning on $K_n(\cdot)$ being constant over time, hence with zero conditional first and second moment. Denote

$$z - \varepsilon_i = (z_1, \ldots, z_{i-1}, z_i - 1, z_{i+1}, \ldots)$$

and assume z has k nonnull components obtained from n particles. Then, as in Section 6, when a particle is removed we have the change of frequency

$$z \mapsto z - \frac{\varepsilon_i}{n} \qquad \text{w.p. } z_i,$$

where $z - \varepsilon_i/n$ has

$$(1): \quad k \text{ nonnull components} \qquad \text{w.p. } 1 - \frac{m_{1,n}}{n},$$

$$(2): \quad k - 1 \text{ nonnull components} \qquad \text{w.p. } \frac{m_{1,n}}{n}.$$

Conditional on case (1), the number of nonnull components remains k if the incoming particle is a copy of an existing type, that is, we observe either of

$$z - \frac{\varepsilon_i}{n} \mapsto z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_i}{n} \qquad \text{w.p. } g_1^{(n,k)}(n_i - 1 - \alpha)/(1 - g_0^{(n,k)}),$$

$$z - \frac{\varepsilon_i}{n} \mapsto z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_j}{n} \qquad \text{w.p. } g_1^{(n,k)}(n_j - \alpha)/(1 - g_0^{(n,k)}),$$

where $g_0^{(n,k)}$ and $g_1^{(n,k)}$ are as in (6.8), while conditional on case (2) we observe

$$z - \frac{\varepsilon_i}{n} \mapsto z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_{k+1}}{n} \qquad \text{w.p. } 1.$$

For $\lambda_n = n^2/2$, the generator of the process $\tilde{Z}^{(n)}(\cdot)$ in this case can be written

$$\mathcal{B}_{n,1} f_1(z)$$

$$= \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \left\{ \lambda_n \delta t \sum_{i=1}^{k} z_i \left[f_1(z) \left(1 - \frac{m_{1,n}}{n} \right) \frac{g_1^{(n,k)}(n_i - 1 - \alpha)}{1 - g_0^{(n,k)}} \right. \right.$$

$$+ \sum_{j \neq i} f_1 \left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_j}{n} \right) \left(1 - \frac{m_{1,n}}{n} \right) \frac{g_1^{(n,k)}(n_j - \alpha)}{1 - g_0^{(n,k)}}$$

$$\left. + f_1 \left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_{k+1}}{n} \right) \frac{m_{1,n}}{n} \right]$$

$$\left. + (1 - \lambda_n \delta t) f_1(z) + O((\delta t)^2) - f_1(z) \right\}$$

for $f_1 \in C^2(\nabla_n)$,

(7.1)
$$\nabla_n = \left\{ z \in \nabla_\infty : z_{n+1} = 0, \sum_{i=1}^{n} z_i = 1 \right\},$$

and ∇_∞ as in (5.8). Exploiting the relation

$$\left(1 - \frac{m_{1,n}}{n} \right) \frac{g_1 \sum_{j=1}^{k} (n_j - \alpha)}{1 - g_0} + \frac{m_{1,n}}{n} = 1,$$

we can write

$$\mathcal{B}_{n,1} f_1(z)$$

$$= \lambda_n \sum_{i=1}^{k} z_i \left\{ [f_1(z) - f_1(z)] \left(1 - \frac{m_{1,n}}{n} \right) \frac{g_1^{(n,k)}(n_i - 1 - \alpha)}{1 - g_0^{(n,k)}} \right.$$

$$+ \sum_{j \neq i} \left[f_1 \left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_j}{n} \right) - f_1(z) \right] \left(1 - \frac{m_{1,n}}{n} \right) \frac{g_1^{(n,k)}(n_j - \alpha)}{1 - g_0^{(n,k)}}$$

(7.2)
$$\left. + \left[f_1 \left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_{k+1}}{n} \right) - f_1(z) \right] \frac{m_{1,n}}{n} \right\}$$

$$= \lambda_n \sum_{i,j} z_i \left[f_1 \left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_j}{n} \right) - f_1(z) \right] \left(1 - \frac{m_{1,n}}{n} \right) \frac{g_1^{(n,k)}(n_j - \alpha)}{1 - g_0^{(n,k)}}$$

$$+ \lambda_n \sum_{i=1}^{k} z_i \left[f_1 \left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_{k+1}}{n} \right) - f_1(z) \right] \frac{m_{1,n}}{n}.$$

By making use of Taylor's theorem, it can be easily verified that the following three relations holds:

$$\sum_{i,j} z_i z_j \left[f_1\left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_j}{n} \right) - f_1(z) \right] = \frac{1}{n^2} \sum_{i,j} z_i (\delta_{ij} - z_j) \frac{\partial^2 f_1}{\partial z_i \partial z_j} + o(n^{-2}),$$

$$\sum_{i,j} z_i \left[f_1\left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_j}{n} \right) - f_1(z) \right] = \frac{1}{n} \sum_i \frac{\partial f_1}{\partial z_i} - \frac{k}{n} \sum_i z_i \frac{\partial f_1}{\partial z_i} + O(n^{-2}),$$

$$\sum_i z_i \left[f_1\left(z - \frac{\varepsilon_i}{n} + \frac{\varepsilon_{k+1}}{n} \right) - f_1(z) \right] = -\frac{1}{n} \sum_i z_i \frac{\partial f_1}{\partial z_i} + O(n^{-2}).$$

By means of the last three expressions, we can write (7.2) as

$$\mathcal{B}_{n,1} f_1(z) = \frac{\lambda_n}{n^2} \sum_{i,j} z_i (\delta_{ij} - z_j) \frac{\partial^2 f_1(z)}{\partial z_i \partial z_j} + \mathcal{B}_{n,1}^{\mathrm{dr}} f_1(z) + o(\lambda_n n^{-2}),$$

where $\mathcal{B}_{n,1}^{\mathrm{dr}} f_1(z)$ is the drift term, given by

$$\mathcal{B}_{n,1}^{\mathrm{dr}} f_1(z) = \frac{\lambda_n}{n} \left[\left(\frac{\alpha k g_1}{1 - g_0} \left(1 - \frac{m_{1,n}}{n} \right) - \frac{m_{1,n}}{n} \right) \sum_i z_i \frac{\partial f_1}{\partial z_i} \right.$$
$$\left. - \frac{\alpha g_1}{1 - g_0} \left(1 - \frac{m_{1,n}}{n} \right) \sum_i \frac{\partial f_1}{\partial z_i} \right].$$

Using (3.5) and Proposition 3.2, it can be seen that

$$\frac{\alpha k g_1}{1 - g_0} \left(1 - \frac{m_{1,n}}{n} \right) - \frac{m_{1,n}}{n} \approx -\frac{\alpha k m_{1,n}}{n^2} \approx -\frac{\alpha^2 s^2}{n}$$

and

$$\frac{\alpha g_1}{1 - g_0} \left(1 - \frac{m_{1,n}}{n} \right) \approx \frac{\alpha}{n},$$

yielding

$$\mathcal{B}_{n,1} f_1(z) = \frac{\lambda_n}{n^2} \sum_{i,j} z_i (\delta_{ij} - z_j) \frac{\partial^2 f_1(z)}{\partial z_i \partial z_j}$$
$$- \frac{\lambda_n}{n^2} \sum_i ((\alpha s)^2 z_i + \alpha) \frac{\partial f_1(z)}{\partial z_i} + o(\lambda_n n^{-2}).$$

For $f \in C(\overline{\nabla}_\infty)$, define $\tilde{r}_n : C(\overline{\nabla}_\infty) \to C(\nabla_n)$ to be

(7.3) $$\tilde{r}_n f = f|_{\nabla_n},$$

namely, the restriction of f to $\tilde{\nabla}_n$, and note that

$$\|\tilde{r}_n f - f\| \longrightarrow 0, \qquad f \in C(\overline{\nabla}_\infty).$$

Recalling that $\lambda_n = n^2/2$ implies that for f as in (5.10) and $\mathcal{L}^{\theta,\alpha}$ as in (1.2),

$$\|\mathcal{L}^{(\alpha s)^2,\alpha} f - \mathcal{B}_{n,1}\tilde{r}_n f\| \to 0.$$

The strong convergence of the corresponding semigroups on $C(\overline{\nabla}_\infty)$, similar to (5.16), and the statement of the proposition now follow from an application of Theorems 1.6.1 and 4.2.11 in Ethier and Kurtz (1986), together with the relativization of the Skorohod topology to $C_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$. $\square$

APPENDIX

Proof of Proposition 3.2.   Consider first that $V_{n,k}$ appearing in (2.5), in the case of generalized gamma processes, can be written [cf. Lijoi, Mena and Prünster (2007b)]

$$V_{n,k} = \frac{a^k}{\Gamma(n)} \int_0^\infty x^n \exp\left\{-\frac{a}{\alpha}[(\tau + x)^\alpha - \tau^\alpha]\right\}(\tau + x)^{\alpha k - n}\,\mathrm{d}x.$$

Together with (2.3), this leads to writing

$$(A.1) \qquad g_0(n,k) = \frac{V_{n,k} - (n - \alpha k)V_{n+1,k}}{V_{n,k}} = 1 - (1 - \alpha k/n)w(n,k),$$

where

$$w(n,k) = \frac{\int_0^\infty x^n \exp\{-(a/\alpha)[(\tau + x)^\alpha - \tau^\alpha]\}(\tau + x)^{\alpha k - n - 1}\,\mathrm{d}x}{\int_0^\infty x^{n-1} \exp\{-(a/\alpha)[(\tau + x)^\alpha - \tau^\alpha]\}(\tau + x)^{\alpha k - n}\,\mathrm{d}x}.$$

Denote by $f(x)$ the integrand of the denominator of $w(n,k)$, so

$$w(n,k) = \int_0^\infty \frac{x}{\tau + x} f(x)\,\mathrm{d}x \Big/ \int_0^\infty f(x)\,\mathrm{d}x.$$

Since $f(x)$ is unimodal, by means of the Laplace method one can approximate $f(x)$ with the kernel of a normal density with mean given by

$$(A.2) \qquad x^* = \arg\max_{x>0} x^{n-1} \exp\left\{-\frac{a}{\alpha}[(\tau + x)^\alpha - \tau^\alpha]\right\}(\tau + x)^{\alpha k - n}$$

and variance given by $-[f''(x)]^{-1}|_{x=x^*}$. It follows that

$$w(n,k) \approx \frac{f_N(x_N^*)C(x_N^*, -[f_N''(x)]^{-1}|_{x=x^*})}{f(x_D^*)C(x_D^*, -[f''(x)]^{-1}|_{x=x^*})},$$

where f_N denotes the integrand of the numerator, x_N^* and x_D^* the modes of the integrands of numerator and denominator, respectively, and $C(x,y)$

is the normalizing constant of a normal kernel with mean x and variance y, yielding

(A.3) $$w(n,k) \approx \frac{f_N(x_N^*)}{f(x_D^*)} \left(\frac{f''(x_D^*)}{f_N''(x_N^*)} \right)^{1/2}.$$

From (A.2), the mode x_D^* is the only positive real root of the equation

(A.4) $$(n-1)x^{-1} + (\alpha k - n)(\tau + x)^{-1} - a(\tau + x)^{\alpha - 1} = 0,$$

which, for $\alpha \neq 1/2, 1/3$, involves finding roots of polynomials of degree greater than 4. When $\alpha = 1/2$ we have

$$
\begin{aligned}
x_D^* = {} & \frac{(k-2)^2}{12a^2} - \frac{\tau}{3} \\
& + \frac{48a^2\tau(n-1)(k-2) + [4a^2\tau - (k-2)^2]^2}{6 \cdot 2^{1/3} a^2 p_{1,D}(a,\tau,n,k)} \\
& + \frac{p_{1,D}(a,\tau,n,k)}{12 \cdot 2^{1/3} a^2},
\end{aligned}
$$

where

$$p_{1,D}(a,\tau,n,k)$$
$$= \{ p_{2,D} + [p_{2,D}^2 + 4(-48a^2\tau(n-1)(k-2) - (4a^2\tau - (k-2)^2)^2)^3]^{1/2} \}^{1/3}$$

and $p_{2,D} = p_{2,D}(a,\tau,n,k)$ with

$$
\begin{aligned}
p_{2,D}(a,\tau,n,k) = {} & 2(k-2)^3[(k-2)^3 - 12a^2\tau(k+4-6n)] \\
& + 96a^4\tau^2(k(k+2) + 10 - 6n(k+4) + 18n^2) \\
& - 128a^6\tau^3.
\end{aligned}
$$

Similarly, one finds that

$$
\begin{aligned}
x_N^* = {} & \frac{(k-2)^2}{12a^2} - \frac{\tau}{3} \\
& + \frac{48a^2\tau n(k-2) + [4a^2\tau - (k-2)^2]^2}{6 \cdot 2^{2/3} a^2 p_{1,N}(a,\tau,n,k)} \\
& + \frac{p_{1,N}(a,\tau,n,k)}{12 \cdot 2^{1/3} a^2},
\end{aligned}
$$

where

$$p_{1,N}(a,\tau,n,k) = [p_{2,N} + (p_{3,N})^{1/2}]^{1/3}$$

with

$$p_{2,N}(a,\tau,n,k) = p_{2,D}(a,\tau,n,k) - 144a^2\tau[4a^2\tau(k+1-6n) - (k-2)^3]$$

and

$$p_{3,N}(a, \tau, n, k)$$
$$= -3^3 2^{10} a^6 n^2 \tau^3 [(k-2)^4 - 2n(k-2)^3$$
$$- 4a^2 \tau [8 + 2k^2 + 9n(3n+4) - 2k(9n+4)] + 16a^4 \tau^2].$$

When $k \approx sn^\alpha$ [cf. (3.7) above] and $\alpha = 1/2$, it can be checked that

(A.5) $$n^{-1} p_{1,i}(a, \tau, n, k) \to 2^{1/3} s^2, \qquad i = N, D,$$

from which

(A.6) $$n^{-1} x_i^* \to (s/2a)^2, \qquad i = N, D.$$

Using this fact, one finds that

$$\frac{f_N(x_N^*)}{f(x_D^*)} \approx \frac{1}{1 + \tau(s/2a)^{-2}}.$$

Computing also the ratio of the two second derivatives, it can be seen that $w(n, k) \to 1$, which by means of (A.1) and (2.6) implies (3.8). In order to find the speed at which $w(n, k)$ goes to 1, consider

$$1 - w(n, k) = \int_0^\infty \frac{\tau}{\tau + x} f(x) \, dx \bigg/ \int_0^\infty f(x) \, dx.$$

The denominator is unchanged, while the mode of $\tau/(\tau + x) f(x)$ is

$$x_N'^* = \frac{(k-4)^2}{12a^2} - \frac{t}{3}$$
$$- \frac{-48a^2 t(n-1)(k-4) - [4a^2 t - (k-4)^2]^2}{6 \cdot 2^{2/3} a^2 p_{1,D}'(a, \tau, n, k)}$$
$$+ \frac{p_{1,D}'(a, \tau, n, k)}{12 \cdot 2^{1/3} a^2},$$

where

$$p_{1,D}'(a, \tau, n, k)$$
$$= \{p_{2,D} + [p_{2,D}^2 + 4(-48a^2 t(n-1)(k-4) - (4a^2 t - (k-4)^2)^2)^3]^{1/2}\}^{1/3}$$

and $p_{2,D} = p_{2,D}(a, \tau, n, k)$ with

$$p_{2,D}'(a, \tau, n, k)$$
$$= 2(k-4)^3 [(k-4)^3 - 12a^2 \tau(k+2-6n)]$$
$$+ 96a^4 \tau^2 (k(k-2) + 10 - 6n(k+2) + 18n^2) - 64a^6 \tau^3.$$

Moreover, $p_{2,D}'$ satisfies (A.5), and (A.6) follows. Unfortunately the fact that the two modes grow with an asymptotically equivalent rate is too rough an

approximation for our purposes here, which ignores how far apart they are if this is negligible with respect to the growth speed. Indeed, it turns out that

$$n^{-1/2}(x_D^* - x_N^*) \to s/a^2.$$

Using this information in the Laplace approximation for $w(n,k) - 1$ yields

$$n(1 - w(n,k)) \to 2a\sqrt{\tau}/s = \beta/s$$

with β as in the statement of the proposition. From (A.1) it is now easy to see that

$$ng_0(n,k) - \alpha k w(n,k) = n(1 - w(n,k)),$$

which provides the second order approximation for $g_0(n,k)$,

$$g_0(n,k) = \frac{\alpha k}{n} + \frac{\beta/s_n}{n} + o(n^{-1}),$$

where the first term is of order $n^{-1/2}$, yielding immediately, by means of (2.6), the second order term for $g_1(n,k)$, that is,

$$g_1(n,k) = \frac{1}{n} - \frac{\beta/s_n}{n^2} + o(n^{-2}).$$

Proof of Theorem 6.1. We can write the generator of (6.10) as

(A.7) $$\mathcal{B}_n(f_0 \times f_1) = f_1 \mathcal{B}_{n,0} f_0 + f_0 \mathcal{B}_{n,1} f_1,$$

where $\mathcal{B}_{n,0}$ and $\mathcal{B}_{n,1}$ drive $K_n(\cdot)/n^\alpha$ and $w(X^{(n)}(\cdot))$, respectively, and $f_0 \in \mathscr{D}(\mathcal{A}_0)$, with $\mathscr{D}(\mathcal{A}_0)$ as in (4.10), $f_1 \in C^2(\nabla_n)$ and ∇_n is as in (7.1). Based on (4.1), while retaining $m_{1,n}$ temporarily, we can write

$$
\begin{aligned}
\mathcal{B}_{n,0} & f_0\left(\frac{k}{n^\alpha}\right) \\
&= \lim_{\delta t \downarrow 0} \frac{1}{\delta t}\left\{ \delta t \left[f_0\left(\frac{k+1}{n^\alpha}\right)\left(1 - \frac{m_{1,n}}{n}\right) g_0^{(n-1,k)} \right.\right. \\
&\qquad\qquad + f_0\left(\frac{k-1}{n^\alpha}\right)\frac{m_{1,n}}{n} g_1^{(n-1,k-1)}(n-1-\alpha(k-1)) \\
&\qquad\qquad + f_0\left(\frac{k}{n^\alpha}\right)\left[1 - \left(1 - \frac{m_{1,n}}{n}\right) g_0^{(n-1,k)} \right. \\
&\qquad\qquad\qquad\qquad \left.\left. - \frac{m_{1,n}}{n} g_1^{(n-1,k-1)}(n-1-\alpha(k-1)) \right]\right] \\
&\qquad\qquad\qquad + (1 - \delta t)f_0\left(\frac{k}{n^\alpha}\right) + O((\delta t)^2) - f_0\left(\frac{k}{n^\alpha}\right)\right\}
\end{aligned}
$$

$$= \left(1 - \frac{m_{1,n}}{n}\right) g_0^{(n-1,k)} \left[f_0\left(\frac{k+1}{n^\alpha}\right) - f_0\left(\frac{k}{n^\alpha}\right)\right]$$

$$+ \frac{m_{1,n}}{n} g_1^{(n-1,k-1)} (n - 1 - \alpha(k-1)) \left[f_0\left(\frac{k-1}{n^\alpha}\right) - f_0\left(\frac{k}{n^\alpha}\right)\right].$$

An application of Taylor's theorem yields

$$\mathcal{B}_{n,0} f_0\left(\frac{k}{n^\alpha}\right)$$

$$= \left(1 - \frac{m_{1,n}}{n}\right) g_0^{(n-1,k)}$$

$$\times \left[f_0\left(\frac{k}{n^\alpha}\right) + \frac{1}{n^\alpha} f_0'\left(\frac{k}{n^\alpha}\right)\right.$$

$$+ \frac{1}{2n^{2\alpha}} f_0''\left(\frac{k}{n^\alpha}\right) + O(n^{-3\alpha}) - f_0\left(\frac{k}{n^\alpha}\right)\Big]$$

$$+ \frac{m_{1,n}}{n} g_1^{(n-1,k-1)} (n - 1 - \alpha(k-1))$$

$$\times \left[f_0\left(\frac{k}{n^\alpha}\right) - \frac{1}{n^\alpha} f_0'\left(\frac{k}{n^\alpha}\right) + \frac{1}{2n^{2\alpha}} f_0''\left(\frac{k}{n^\alpha}\right)\right.$$

$$+ O(n^{-3\alpha}) - f_0\left(\frac{k}{n^\alpha}\right)\Big]$$

$$= \frac{1}{n^\alpha} f_0'\left(\frac{k}{n^\alpha}\right) \left[\left(1 - \frac{m_{1,n}}{n}\right) g_0^{(n-1,k)}\right.$$

$$- \frac{m_{1,n}}{n} g_1^{(n-1,k-1)} (n - 1 - \alpha(k-1))\Big]$$

$$+ \frac{1}{2n^{2\alpha}} f_0''\left(\frac{k}{n^\alpha}\right) \left[\left(1 - \frac{m_{1,n}}{n}\right) g_0^{(n-1,k)}\right.$$

$$+ \frac{m_{1,n}}{n} g_1^{(n-1,k-1)} (n - 1 - \alpha(k-1))\Big]$$

$$+ O(n^{-4\alpha}).$$

From (4.4) and (4.5) we have

$$\mathcal{B}_{n,0} f_0\left(\frac{k}{n^\alpha}\right) = \frac{1}{n^\alpha} f_0'\left(\frac{k}{n^\alpha}\right) \left[\frac{\beta/s_n}{n} + o(n^{-1})\right]$$

$$+ \frac{1}{2n^{2\alpha}} f_0''\left(\frac{k}{n^\alpha}\right) \left[\frac{s_n}{n^\alpha} + O(n^{-1-2\alpha})\right] + O(n^{-4\alpha})$$

$$= \frac{\beta/s_n}{n^{1+\alpha}} f_0'\left(\frac{k}{n^\alpha}\right) + \frac{s_n}{2n^{3\alpha}} f_0''\left(\frac{k}{n^\alpha}\right) + o(n^{-1-\alpha}),$$

from which it follows that

$$\text{(A.8)} \qquad \|\mathcal{A}_0 f_0 - n^{3/2} \mathcal{B}_{n,0} f_0\| \to 0$$

as $n \to \infty$, with $\mathcal{A}_0$ as in (4.9). Since (6.11) implies $K_n(0)/n^\alpha \Rightarrow S_0$, the previous expression, together with Theorems 1.6.1 and 4.2.5 in Ethier and Kurtz (1986), implies (4.3) with $C_{[0,\infty)}([0,\infty))$ replaced by $D_{[0,\infty)}([0,\infty))$, while (4.3) follows from relativization of the Skorohod topology to $C_{[0,\infty)}([0,\infty))$.

In order to describe $\mathcal{B}_{n,1}$ in (A.7), define

$$\text{(A.9)} \quad \phi_n(\mu) = f_1(\langle h_1, \mu \rangle, \ldots, \langle h_n, \mu \rangle), \qquad f_1 \in C_0^2(\mathbb{R}^n), h_i \in C(\mathbb{X}),$$

and

$$\text{(A.10)} \qquad \mu_n(t) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(t)}, \qquad t \geq 0.$$

Then the generator of the $\mathscr{P}(\mathbb{X})$-valued process $\mu_n(\cdot) = \{\mu_n(t), t \geq 0\}$ can be written

$$
\begin{aligned}
\mathbb{B}_n \phi_n(\mu) &= \frac{1}{n} \sum_{i=1}^n g_0^{(n-1,i)}[\langle Ph_i, \mu \rangle - \langle h_i, \mu \rangle] \frac{\partial f_1}{\partial z_i} \\
\text{(A.11)} \qquad &- \frac{\alpha}{n} \sum_{i=1}^n g_1^{(n-1,i)} K_{n-1,i}[\langle Q^{(n-1,i)} h_i, \mu \rangle - \langle h_i, \mu \rangle] \frac{\partial f_1}{\partial z_i} \\
&+ \frac{1}{n} \sum_{1 \leq k \neq i \leq n} g_1^{(n-1,i)}[\langle h_i h_j, \mu \rangle - \langle h_i, \mu \rangle \langle h_j, \mu \rangle] \frac{\partial^2 f_1}{\partial z_i \partial z_j},
\end{aligned}
$$

where g_{z_i} is the derivative of g with respect to its ith argument and $Q^{(n-1,i)}$ is defined as

$$\text{(A.12)} \qquad Q^{(n-1,i)} g(z) = \int g(y) p_{n-1,i}^*(z, \mathrm{d}y), \qquad g \in C_0(\mathbb{X}),$$

with

$$\text{(A.13)} \qquad p_{n-1,i}^*(z, \mathrm{d}y) = \frac{1}{K_{n-1,i}} \sum_{j=1}^{K_{n-1,i}} \delta_{x_j^*}(\mathrm{d}y).$$

Unlabel now the model by choosing $\phi_n(\mu)$ as in (A.9) with $h_j(\cdot)$ being the indicator function of the jth largest atom in μ, so that $\langle h_j, \mu \rangle = z_j$ is the relative frequency associated with the jth most abundant species. Note that some arguments of $f_1(z_1, \ldots, z_n)$ can be null since $K_{n-1,i} \leq K_n \leq n$. With this choice we have $\langle h_i h_j, \mu \rangle - \langle h_i, \mu \rangle \langle h_j, \mu \rangle = z_i \delta_{ij} - z_i z_j$, where δ_{ij} is the

Knonecker delta, and, under (6.6) and (A.12),

$$\langle Ph_i, \mu \rangle - \langle h_i, \mu \rangle = \sum_{j=1}^{n} p_1(x_j^*, \mathrm{d}x_i^*)z_j - z_i$$

$$= \sum_{j=1}^{n} [p_1(x_j^*, \mathrm{d}x_i^*) - \delta_{ij}]z_j$$

and

$$\langle Q^{(n-1,i)}h_i, \mu \rangle - \langle h_i, \mu \rangle = \sum_{j=1}^{n} [p_{n-1,i}^*(x_j^*, \mathrm{d}x_i^*) - \delta_{ij}]z_j.$$

It follows that $\mathbb{B}_n$ reduces to

$$\begin{aligned}
\mathcal{B}_{n,1}f_1 = \; & \frac{1}{n}\sum_{i=1}^{n} g_0^{(n-1,i)}\left(\sum_{j=1}^{n}[p_1(x_j^*, \mathrm{d}x_i^*) - \delta_{ij}]z_j\right)\frac{\partial f_1}{\partial z_i} \\
& - \frac{\alpha}{n}\sum_{i=1}^{n} g_1^{(n-1,i)}K_{n-1,i}\left(\sum_{j=1}^{n}[p_{n-1,i}^*(x_j^*, \mathrm{d}x_i^*) - \delta_{ij}]z_j\right)\frac{\partial f_1}{\partial z_i} \\
& + \frac{1}{n}\sum_{i,j=1}^{n} g_1^{(n-1,i)}z_i(\delta_{ij} - z_j)\frac{\partial^2 f_1}{\partial z_i \partial z_j}.
\end{aligned}$$

(A.14)

Here a mutation from type i to type j occurs at rate $q_{ij} = q_{ij}(z)$ given by

$$\begin{aligned}
q_{ij} = \; & \frac{1}{n}[g_0^{(n-1,i)}[p_1(i, \{j\}) - \delta_{ij}] \\
& - \alpha g_1^{(n-1,i)}K_{n-1,i}[p_{n-1,i}^*(i, \{j\}) - \delta_{ij}]],
\end{aligned}$$

(A.15)

where $p_1(i, \{j\})$ and $p_{n-1,i}^*(i, \{j\})$ stand for $p_1(x_i^*, \mathrm{d}x_j^*)$ and $p_{n-1,i}^*(x_i^*, \mathrm{d}x_j^*)$. When (6.7) holds, from the nonatomicity of ν_0 we have $p_1(z, \mathrm{d}y) = 0$ for every $y \in \mathbb{X}$, and when (A.13) holds, we have $p_{n-1,i}^*(z, \mathrm{d}y) = K_{n-1,i}^{-1}$, from which

$$\begin{aligned}
q_{ij} &= \frac{1}{n}[-\delta_{ij}g_0^{(n-1,i)} - \alpha g_1^{(n-1,i)} + \alpha g_1^{(n-1,i)}K_{n-1,i}\delta_{ij}] \\
&= \frac{1}{n}[-\delta_{ij}(1 - (n-1)g_1^{(n-1,i)}) - \alpha g_1^{(n-1,i)}] \\
&= \frac{1}{n}\left[-\delta_{ij}\left(1 - (n-1)\left(\frac{1}{n-1} - \frac{\beta/s}{(n-1)^2}\right)\right)\right. \\
&\qquad \left. - \alpha\left(\frac{1}{n-1} - \frac{\beta/s}{(n-1)^2}\right)\right] + o(n^{-2}) \\
&= -\frac{\delta_{ij}\beta/s}{n(n-1)} - \frac{\alpha}{n(n-1)} + o(n^{-2}),
\end{aligned}$$

(A.16)

where the second equality follows from (2.6) and the third from (3.9). Once again it is clear that the key point for determining the limiting behavior of the diffusion is the second order approximation of the predictive weights, as obtained in Proposition 3.2. Hence, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}z_j = -\sum_{i=1}^{n}\left[\frac{z_i\beta/s}{n(n-1)} + \frac{\alpha}{n(n-1)} + o(n^{-2})\right],$$

from which (A.14), substituting (3.9) in the third term, reduces to

$$\mathcal{B}_{n,1}f_1 = \sum_{i,j=1}^{n}[n^{-2} - O(n^{-3})]z_i(\delta_{ij} - z_j)\frac{\partial^2 f_1}{\partial z_i \partial z_j}$$

$$- \sum_{i=1}^{n}\left[\frac{z_i\beta/s}{n(n-1)} + \frac{\alpha}{n(n-1)} + o(n^{-2})\right]\frac{\partial f_1}{\partial z_i},$$

which in turn implies that

$$(A.17) \qquad \|\mathcal{A}_1 f_1 - (n^2/2)\mathcal{B}_{n,1}\tilde{r}_n f_1\| \to 0$$

with $\mathcal{A}_1$ as in (5.9), $f_1 \in \mathscr{D}(\mathcal{A}_1)$ as in (5.10) and $\tilde{r}_n$ as in (7.3). From (A.8) and (A.17) it follows that

$$\|\mathcal{A}(f_0 \times f_1) - (\tilde{r}_n f_1)n^{3/2}\mathcal{B}_{n,0}f_0 - f_0 n^2 \mathcal{B}_{n,1}\tilde{r}_n f_1/2\| \to 0$$

with $\mathcal{A}$ as in (1.4) and $(f_0 \times f_1)$ as in (5.11). The fact that (6.12) holds in $D_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ follows from the density of (5.11) in $C_0([0,\infty) \times \overline{\nabla}_\infty)$, together with Theorems 1.6.1 and 4.2.11 in Ethier and Kurtz (1986), which imply, respectively, the strong convergence of the associated semigroups on $C_0([0,\infty) \times \overline{\nabla}_\infty)$, similarly to (5.16), and the weak convergence of the probability measures induced on $D_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$. The same assertion with $C_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ in place of $D_{[0,\infty)\times\overline{\nabla}_\infty}([0,\infty))$ now follows from relativization of the Skorohod topology.

REFERENCES

BERESTYCKI, N. and PITMAN, J. (2007). Gibbs distributions for random partitions generated by a fragmentation process. *J. Stat. Phys.* **127** 381–418. MR2314353

BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge Univ. Press, Cambridge. MR2253162

BERTOIN, J. (2008). Two-parameter Poisson–Dirichlet measures and reversible exchangeable fragmentation-coalescence processes. *Combin. Probab. Comput.* **17** 329–337. MR2410390

BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614

BRIX, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Probab.* **31** 929–953. MR1747450

CHARALAMBIDES, C. A. (2005). *Combinatorial Methods in Discrete Distributions*. Wiley, Hoboken, NJ. MR2131068

ETHIER, S. N. and KURTZ, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.* **13** 429–452. MR0615945

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York. MR0838085

FAVARO, S., LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2009). Bayesian non-parametric inference for species variety with a two-parameter Poisson–Dirichlet process prior. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 993–1008. MR2750254

FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics Models and Asymptotic Behaviors*. Springer, Heidelberg. MR2663265

FENG, S. and SUN, W. (2010). Some diffusion processes associated with two parameter Poisson–Dirichlet distribution and Dirichlet process. *Probab. Theory Related Fields* **148** 501–525. MR2678897

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

GNEDIN, A. (2010). A species sampling model with finitely many types. *Electron. Commun. Probab.* **15** 79–88. MR2606505

GNEDIN, A. and PITMAN, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **325** 83–102, 244–245. MR2160320

GOLDSCHMIDT, C., MARTIN, J. B. and SPANÒ, D. (2008). Fragmenting random permutations. *Electron. Commun. Probab.* **13** 461–474. MR2430713

GRIFFITHS, R. C. and SPANÒ, D. (2007). Record indices and age-ordered frequencies in exchangeable Gibbs partitions. *Electron. J. Probab.* **12** 1101–1130. MR2336601

HANDA, K. (2009). The two-parameter Poisson–Dirichlet point process. *Bernoulli* **15** 1082–1116. MR2597584

HO, M. W., JAMES, L. F. and LAU, J. W. (2007). Gibbs partitions (EPPF's) derived from a stable subordinator are Fox H- and Meijer G-transforms. Available at arXiv:0708.0619v2.

KARLIN, S. and TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York. MR0611513

KINGMAN, J. F. C. (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B* **37** 1–22. MR0368264

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291. MR2236441

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007a). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94** 769–786. MR2416792

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007b). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 715–740. MR2370077

Lijoi, A., Prünster, I. and Walker, S. G. (2008a). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18** 1519–1547. MR2434179

Lijoi, A., Prünster, I. and Walker, S. G. (2008b). Investigating nonparametric priors with Gibbs structure. *Statist. Sinica* **18** 1653–1668. MR2469329

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker, eds.) 80–136. Cambridge Univ. Press, Cambridge. MR2730661

McCullagh, P., Pitman, J. and Winkel, M. (2008). Gibbs fragmentation trees. *Bernoulli* **14** 988–1002. MR2543583

Petrov, L. A. (2009). A two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funct. Anal. Appl.* **43** 279–296.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249

Pitman, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory* (T. S. Ferguson, L. S. Shapley and J. B. MacQueen, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. IMS, Hayward, CA. MR1481784

Pitman, J. (2003). Poisson–Kingman partitions. In *Statistics and Science: A Festschrift for Terry Speed* (D. R. Goldstein, ed.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. IMS, Beachwood, OH. MR2004330

Pitman, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

Pitman, J. and Yor, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129

Ruggiero, M. and Walker, S. G. (2009). Countable representation for infinite dimensional diffusions derived from the two-parameter Poisson–Dirichlet process. *Electron. Commun. Probab.* **14** 501–517. MR2564485

Teh, Y. W. and Jordan, M. I. (2010). Bayesian nonparametrics in machine learning. In *Bayesian Nonparametrics* (N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker, eds.). Cambridge Univ. Press, Cambridge.

M. Ruggiero
S. Favaro
Department of Economics
  and Statistics
University of Torino
Corso Unione Sovietica 218/bis
10134, Torino
Italy
E-mail: matteo.ruggiero@unito.it
       stefano.favaro@unito.it

S. G. Walker
Institute of Mathematics, Statistics
  and Actuarial Science
University of Kent
CT2 7NZ, Canterbury
United Kingdom
E-mail: S.G.Walker@kent.ac.uk